The use of reflectance classification for chlorophyll algorithm application across multiple optical water types in South African coastal waters

Marié Smith

Thesis presented for the degree of
Doctor of Philosophy

Oceanography Department
University of Cape Town
August 2016

Declaration

The present work has been originally written by me, with the full support of my supervisors: Dr Marcello Vichi of the Department of Oceanography, University of Cape Town, South Africa; Dr Stewart Bernard of the Council for Scientific and Industrial Research, South Africa; Dr Mark Matthews of CyanoLakes (Pty) Ltd, Cape Town, South Africa. In addition, important contributions to the approach used in this study are clearly acknowledged through referencing within the text.

Abstract

Ocean colour remote sensing is a valuable tool for deriving information about key biogeochemical variables over inland, coastal and ocean waters at scales unachievable via *in situ* techniques. However, broader use of ocean colour data is still limited by the need for users to choose among a seemingly complicated range of available satellite products and to understand the limitations and constraints of these products across a wide range of water types. This issue could benefit from the capability to seamlessly apply and blend water-type appropriate algorithms into a single output product that provides optimal retrievals over a wide range of water types. The assessment of the fuzzy membership of satellite remote sensing reflectance (R_{rs}) to pre-defined regional optical water types (OWTs) provides a framework for application and blending of OWT-appropriate algorithms on a per-pixel basis. This study presents the first characterization of the OWTs in the coastal waters of South Africa. The OWTs are determined through stepwise fuzzy c-means clustering of a systematically expanding and modified database constructed from *in situ*, synthetic and regionally extracted Medium Resolution Imaging Spectrometer (MERIS) R_{rs}. A database division allows separate and more detailed clustering of phytoplankton-dominated R_{rs} and backscattering-dominated R_{rs} into six and five classes respectively. Chlorophyll *a* (Chl *a*) algorithms are assigned per OWT based on lowest error and uncertainty. The blended Chl *a* product consists of weighted retrievals from five different algorithms, including two 4^{th} order polynomial exponential algorithms utilizing the blue-green spectral region, two red-NIR band ratio algorithms, and a neural network. The algorithm blending procedure retrieves satellite-derived Chl *a* concentration ([Chl *a*]) with lower RMS error and uncertainty compared to individual algorithms and provides improved capability to retrieve [Chl *a*] for different South African water types with a single product over a range spanning almost four orders of magnitude. The eleven OWTs are utilized in the classification and algorithm blending framework and applied to the full archive of MERIS Level 2 reflectance between the years 2002 and 2012 over South Africa's coastal waters. The persistence of the OWTs is presented and linked to the prominent environmental and physical drivers, whilst regions with low total class membership sums are discussed

in terms of satellite data coverage and data quality. A time series of the blended [Chl *a*] product displays improved capability to capture the ranges of variability observed in the coastal, shelf and offshore environment compared to currently available regional and standard MERIS Level 2 products.

Acknowledgements

For providing funding and data, I would like to thank the following:
ACCESS for three and a half years of funding, covering my living costs, fieldwork and enabling me to attend numerous conferences. ESA for providing the satellite data needed for the project. ACRI-ST, ARGANS and ESA for access to the MERMAID system, and the following individual contributions to these data: Vanda Brotas and Carolina Sà for PortCoast data; David McKee for BristolIrishSea data; David Antoine for BOUSSOLE data; Sélima Mustapha and Simon Belanger for CASES data; Catherine Belin for Ifremer-REPHY data; David Siegel for PlumesAndBlooms data; Antonio Mannino, Frank Muller-Karger, Rick Gould, Kendall Carder, Greg Mitchell, David Siegel for NOMAD data; Guiseppe Zibordi, Jean-francois Berthon, and Elisabetta Canuti for BioOptEuroFleets data; Hubert Loisel for EastEngChannel and FrenchGuiana data. CSIR and DAFF for additional funding, equipment and research expertise needed for fieldwork.

For everything else, I would like to thank the following individuals:
Stewart Bernard for project supervision, advice, financial support, endless input on chapters, and keeping me relatively sane. Marcello Vichi and Mark Matthews for providing valuable input and feedback on my chapters. Hayley Evers-King and Lisl Robertson-Lain for limitless technical support and being my sounding board. Andy Rabagliati, Raymond Roman and Luke Gregor for technical support and assistance with data processing. My fellow PhD students Ffion Atkins, Lauren Biermann, Emma Bone and Christo Whittle for sharing and empathizing with my frustrations. I would like to thank my parents for inspiring my love for science and the ocean, and for continued emotional and financial support throughout my graduate and post-graduate career, enabling me to focus solely on my work. For all the laughs, wine, surf sessions and electronic rants, I want to thank my two best friends Karien Brand and William Dowd. Final thanks goes to my partner Rod Smith, who has provided me with all the love, support, and ice cream that a girl could ever need to get through this PhD!

List of notation

a	Total absorption coefficient (m^{-1})
a_ϕ	Phytoplankton absorption coefficient (m^{-1})
a_{dg}	Combined CDOM and detritus absorption coefficient (m^{-1})
a_g	Absorption coefficient of CDOM (m^{-1})
a_{NAP}	Absorption coefficient of non-algal particulate matter (m^{-1})
a_w	Absorption coefficient of seawater (m^{-1})
A	Average total class membership (per image)
AOPs	Apparent optical properties
$b_{b\phi}$	Phytoplankton backscattering coefficient (m^{-1})
$b_{b\phi}/b_\phi$	Phytoplankton backscatter fraction
b_{bs}	Non-algal particulate backscattering coefficient (m^{-1})
[Chl a]	Chlorophyll a concentration (mg m^{-3})
CDOM	Coloured dissolved organic matter
CE	Classification entropy
CZCS	Coastal zone color scanner
E_d	Downwelling irradiance (μW cm^{-2} nm^{-1})
EAP	Equivalent algal population
EOF	Empirical orthogonal function
FCM	Fuzzy c-means
f_i	Membership function
G2B	Gilerson et al. (2010) 2-band algorithm
G3B	Gilerson et al. (2010) 3-band algorithm
TSRB	Satlantic hyperspectral tethered surface radiometer buoy
IOPs	Inherent optical properties
IOCCG	International ocean colour coordinating group
IR	Infrared
L_u	Upwelling radiance (μW cm^{-2} nm^{-1} sr^{-1})
L_w	Water-leaving radiance (μW cm^{-2} nm^{-1} sr^{-1})

KZN	KwaZulu-Natal
MARE	Median absolute relative error
MERCI	MERIS catalogue and inventory
MERIS	Medium resolution imaging spectrometer
MERMAID	MERIS matchup in-situ database
MODIS	Moderate resolution imaging spectroradiometer
MPH	Maximum peak-height
NAP	Non-algal particulate matter
NIR	Near infrared
NOMAD	NASA bio-optical marine algorithm dataset
OC3E	3-band algorithm developed for MODIS (O'Reilly et al., 1998), with coefficients derived for MERIS wavelengths
OC4E	4-band algorithm developed for SeaWiFS (O'Reilly et al., 1998), with coefficients derived for MERIS wavelengths
OCMe	MERIS algal pigment index for Case 1 waters algorithm (Morel and Antoine, 2011)
OLCI	Ocean and land colour instrument
OWTs	Optical water types
PC	Partition coefficient
PCA	Principal component analysis
PCD1_13	Quality flag for the 13 MERIS reflectance bands
RMSE	Root-mean-square error
R_n	Integral-normalized R_{rs}
RR	Reduced resolution
R_{rs}	Remote sensing reflectance (sr^{-1})
ρ_w	Water-leaving reflectance
S	Separation index
SC	Partition index
SeaWiFS	Sea-viewing wide field-of-view sensor
TOA	Top of atmosphere
z	Depth (m)
Z_i^2	Squared Mahalanobis distance

Contents

Chapter 1

Introduction

1.1 General introduction

The ocean can be conceptualized as a collection of dynamic biogeochemical or ecological provinces resulting from various physical forcing mechanisms (IOCCG, 2009a). These provinces often exhibit different ranges of bio-optical properties which give rise to identifiable changes in the shape and magnitude of the water-leaving reflectance spectra across regions. Provinces with similar bio-optical properties are often called *optical water types* (OWTs). Phytoplankton, the base of aquatic food webs, is one of the primary optically significant constituents in the water column. Quantitative observations of phytoplankton biomass are necessary for ecosystem, aquaculture and fisheries management, and provide useful indices for water quality and climate monitoring. The need for synoptic views of spatial and temporal phytoplankton variability has been a driving force for the advancement within the field of ocean colour remote sensing over the last few decades. Default satellite products generally only utilize a single algorithm over regions potentially containing many different OWTs; however different OWTs may require discrete algorithm considerations to derive accurate quantitative information about the biogeochemical constituents in the water column. Many end-users of ocean colour products are non-specialists who may not be aware of the algorithm limitations and resultant uncertainties inherent to these satellite products. Resultantly, the ocean colour community has identified the need for a framework that can apply the most appropriate algorithm per water type in order to facilitate the lowest errors and uncertainty across the image, and also seamlessly blend the outputs together into a single product (IOCCG, 2009a). The OWT approach to algorithm application offers a dynamic scaling capability by classifying water-leaving reflectance spectra and applying optimal algorithms on a per-pixel basis across a wide range of bio-optical conditions and phytoplankton biomass concentrations. This provides

an avenue to a more globally applicable approach to ocean colour, something that is most likely unattainable with a single algorithm.

1.1.1 The marine light field and the role of inherent optical properties

In the third report from the International Ocean-Colour Coordinating Group (IOCCG, 2000, p.26) the goal of remote sensing of ocean colour was specified as: "to derive quantitative information on the types of substances present in the water and on their concentrations, from variations in the spectral form and magnitude of the ocean-colour signal". Since satellites cannot measure the in-water constituents or their concentrations directly, it is necessary to cultivate an understanding of how these substances interact with the

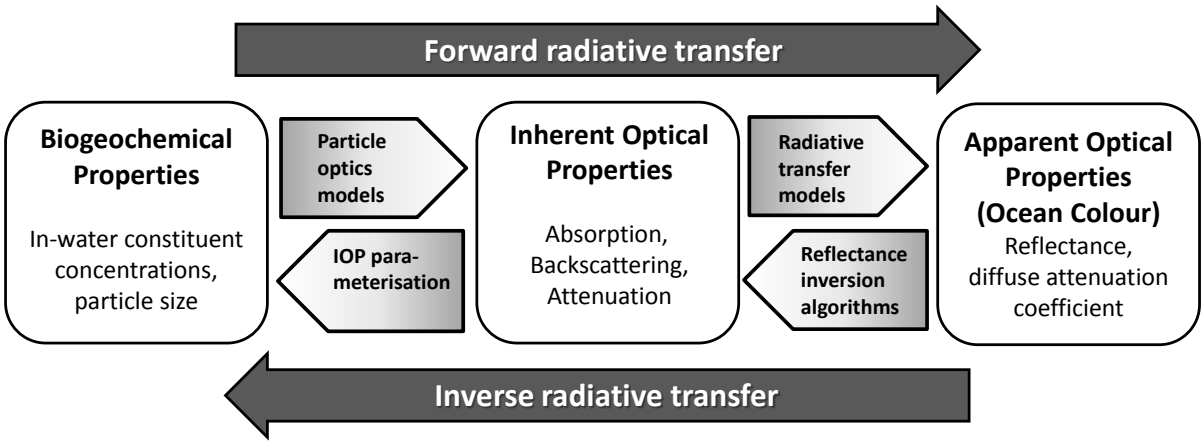


Figure 1.1: Schematic outlining the relationships between the apparent optical properties, inherent optical properties and biogeochemical properties as defined by radiative transfer theory.

Photons of light within the water can either be absorbed or scattered. The extent to which the photons are scattered and absorbed depends on the intrinsic characteristics of the water as well as the substances within; these are referred to as the inherent optical properties (IOPs) (Preisendorfer, 1976) since they only depend on the type and concentrations of the substances in the water and not on the geometric structure of the marine light field. The full suite of IOPs can be decomposed and parametrized according to the specific primary optically significant constituents that affect the ocean colour, i.e. seawater,

phytoplankton, coloured dissolved organic material (CDOM), and non-algal particulate matter (NAP). For instance the total absorption coefficient ($a(\lambda)$) can be expressed as follows:

$$a(\lambda) = a_w(\lambda) + a_\phi(\lambda) + a_g(\lambda) + a_{NAP}(\lambda) \tag{1.1}$$

where a_w, a_ϕ, a_g, and a_{NAP} indicates the absorption coefficients of seawater, phytoplankton, CDOM and NAP, respectively. Other water constituents such as bubbles, viruses, bacteria and colloids (Stramski et al., 2004) may also contribute to the total IOP budget; however these will not specifically be addressed in this study. The apparent optical properties (AOPs) are influenced by the incident light field, whilst also being affected by the substances in the medium. AOPs are traditionally derived from measurements made by radiometers, such as downwelling irradiance (E_d) and upwelling radiance (L_u); although these two variables are not defined as AOPs, they can be used in ratios which in turn form stable properties and display regular features which are useful for describing a water body (Mobley, 2010). A widely used AOP in the field of ocean colour remote sensing and ocean optics is the remote sensing reflectance (R_{rs}), as defined in equation 1.2.

Forward approaches are used to estimate the ocean colour from the IOPs by applying radiative transfer methods such as reflectance approximation techniques (Zaneveld, 1995) or more exact models such as Hydrolight (Mobley, 2011). The inverse method solves the IOPs and optically significant constituents from *in situ* or satellite-based radiometric R_{rs} measurements with the use of various approaches including non-linear optimisation techniques, principal component analysis, neural networks, and algebraic approaches such as semi-analytical algorithms (IOCCG, 2000, 2006). Since most ocean colour sensors cannot measure the directional distribution of the marine light field, various assumptions are introduced in the inversion methodology. Many semi-analytical algorithms use the reflectance approximation as a framework to incorporate IOPs and light field parameters when inverting R_{rs}, as seen in equation 1.2:

$$R_{rs} = \frac{L_w(0^+, \lambda)}{E_d(0^+, \lambda)} = \frac{f}{Q}\eta^2\tau\frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \tag{1.2}$$

where $L_w(0^+, \lambda)$ is the water-leaving radiance just above the water surface; $E_d(0^+, \lambda)$ is the downwelling irradiace just above the water surface; f/Q is a shape factor describing the angular structure of the light field as influenced by the IOPs, solar zenith angle and surface roughness (Morel et al., 2002); $\eta^2\tau$ are the transmission parameters for the air/sea interface; while $a(\lambda)$ and $b_b(\lambda)$ are the total absorption and backscattering coefficients

respectively. This inversion of R_{rs} can often produce ambiguous results, since many IOP combinations may produce identical ocean colour spectra (Defoin-Platel and Chami, 2007).

Many inversion algorithms are aimed at determining chlorophyll *a* concentration ([Chl *a*]), as it is a parameter often used as a proxy for phytoplankton biomass. Chl *a* is related, in a non-linear fashion, to algal IOPs and is only indirectly (if at all) related to the other IOPs that make up an R_{rs} signal (IOCCG, 2006). Many constituent-specific IOP parametrizations are regionally or seasonally defined which can limit their scale of application (e.g. Cota et al., 2003; Smyth et al., 2006; Tzortziou et al., 2006; Ambarwulan et al., 2011; Tilstone et al., 2012). It is thus vital to apply regionally or water type appropriate algorithms in order to retrieve optimal quantitative estimates of [Chl *a*]. A wide variety of techniques exist to derive [Chl *a*] in waters with varying degrees of other optically significant constituents; these algorithms generally have optimal ranges of retrieval and operational constraints, some of which will be discussed below.

1.1.2 Chlorophyll *a* algorithms

Approaches commonly used for low biomass phytoplankton dominated waters

Chlorophyll *a* concentration ([Chl *a*]) has long been used as a proxy for phytoplankton biomass, as it is the primary photosynthetic pigment found in all phytoplankton (O'Reilly et al., 1998). Phytoplankton have strong absorbing qualities in the blue and red spectral regions, which allow [Chl *a*] to be related to water-leaving reflectance either empirically or analytically. Historically, some of the first bio-optical algorithms were based on the assumption that phytoplankton covary with their degradation material in the water; these regions were known as Case 1 waters (Morel and Prieur, 1977) and often found in the open ocean areas where phytoplankton are the primary optically significant constituents (Morel and Prieur, 1977; Morel, 1980).

At low pigment concentration ranges (< 0.25 mg m^{-3} Chl *a*) found in about 78 % of the global ocean (Hu et al., 2012) the red chlorophyll band has very little influence on the ocean colour (Yentsch, 1960); in these areas the blue-green spectral region represents the strongest signal and shows the best relationship with water-leaving reflectance bands. The most common algorithms designed for ocean colour sensors have used relatively simple empirical relationships between the blue-green spectral bands to determine [Chl *a*] in Case 1 waters, such as OC3 and OC4 for MODIS and SeaWiFS respectively

(O'Reilly et al., 1998) and the algal pigment index 1 (Algal1) for MERIS (Morel and Antoine, 2011); the generalized regression coefficients of these algorithms are usually derived from global datasets such as the SeaWiFS Bio-optical Algorithm Mini-workshop (SeaBAM) dataset (Firestone and Hooker, 1998) and the NASA bio-Optical Marine Algorithm Dataset (NOMAD) (Werdell and Bailey, 2005). These types of algorithms use the "black box" approach (IOCCG, 2006) where the R_{rs} are related directly to the [Chl a], effectively bypassing the IOPs. Although they are generally globally and seasonally robust, many of these standard globally derived ocean colour algorithms have shown variable performance at regional (e.g. Volpe et al., 2007) and ocean basin (e.g. Szeto et al., 2011) scales, which has primarily been attributed to challenges in satellite atmospheric correction and differences in local IOPs (McClain, 2009; Sauer et al., 2012). Some studies account for this variability by using regional validation data to re-parametrise the regression coefficients (e.g. McKee et al., 2007a; Volpe et al., 2007; Pan et al., 2008) or adjusted band combinations (e.g. Mitchell and Kahru, 2009) of the standard empirical algorithms for optimal retrievals in their local waters. There are also a range of other approaches for Case 1 waters beyond the empirical blue-green band ratios, some of which have included spectral band difference algorithms such as the fluorescence line height (FLH; Letelier and Abbott, 1996; Gower et al., 1999) and the colour index algorithm (CIA; Hu et al., 2012), and semi-analytical algorithms (e.g. Garver and Siegel, 1997; Hoge et al., 1999), among others.

Most Chl a algorithms are designed to operate in optically deep waters, and could produce severe overestimations in shallow waters (D'Sa et al., 2002; Schaeffer et al., 2012) where light reflected from the bottom can contribute to the water-leaving reflectance signal; in these cases special algorithms or corrections may be required to resolve the optically significant constituents (e.g. Lee et al., 1999; Cannizzaro and Carder, 2006b). The assumptions and optical relationships that many Case 1 algorithms are based on can break down when other optically significant constituents affect the visible light spectrum. Such conditions may often occur in coastal and inland waters where suspended sediments or CDOM vary independently of phytoplankton in the water column; these have often been referred to as Case 2 waters (Morel and Prieur, 1977). The blue-green ratio algorithms may also fail in waters with high concentrations of phytoplankton biomass, since the high amounts of phytoplankton absorption can invalidate the band-ratio relationships between [Chl a] and the blue-green spectral region. The fluorescence signal is also affected strongly by other water constituents in optically complex waters; studies have demonstrated that the FLH signal could be masked by non-algal materials in turbid waters (McKee et al.,

2007b; Gilerson et al., 2008), while the algorithm can fail at [Chl a] > 20 mg m^{-3} (Gower and King, 2007). Algorithms utilising different spectral regions or analytical approaches are often required to determine phytoplankton biomass concentrations in these optically complex waters.

Approaches commonly used in optically complex waters

Under eutrophic or high phytoplankton biomass conditions algorithms centred around the red-NIR wavelengths have been preferred to the blue-green spectral region. These algorithms operate on the red-edge, which represents the changes in the position and magnitude of the Chl a fluorescence and particulate backscatter and absorption related peaks (Matthews et al., 2012), and have included various band-ratio algorithms (e.g. Koponen et al., 2007; Moses et al., 2009a,b; Matthews et al., 2010), as well as spectral band difference algorithms such the maximum chlorophyll index (MCI; Gower et al., 2005) and maximum peak-height algorithms (MPH; Matthews et al., 2012). The use of the red spectral region is advantageous when using atmospherically corrected R_{rs} spectra, since the potential aerosol extrapolation error from the NIR to the red bands is generally less extreme than to the green or blue spectral region (McClain, 2009); the MPH has circumvented the aerosol correction problem altogether by applying the algorithm to Rayleigh-corrected reflectance (Matthews et al., 2012).

Model-based algorithms are often preferred in optically complex waters. These algorithms apply inversion techniques to the satellite derived or *in situ* measurements of R_{rs} to obtain IOPs or concentrations of water constituents; they have the advantage that several ocean colour properties can be derived simultaneously (Maritorena et al., 2002). Inversion schemes have included semi-analytical models (Carder et al., 1999; Lee et al., 2002; Maritorena et al., 2002), linear inversion based on principal component analysis (Krawczyk et al., 2004) and neural networks (Doerffer and Schiller, 2007; El-habashi and Ahmed, 2015), among others. The development of these approaches often relies on the collection of *in situ* data, knowledge of regional IOP variability and specificity, as well as the technical and statistical complexities involved in training and optimizing a forward model. As a result the successful utility of these algorithms is inherently dependent on, and limited by, the ranges of their training data or derivation, and the algorithms are often only regionally applicable for certain water types (e.g. D'Sa et al., 2006; Matsuoka et al., 2013; Zhu and Yu, 2013). Over the past few years software has been developed to aid with the development and testing of IOP models; an example is the Generalized IOP (GIOP) model (Franz and Werdell, 2010) which provides a framework with a variety of

published parameterizations for the evaluation, construction and regional tuning of IOP models. Adequate performance of these inversion models can often only be achieved with accurate water-leaving reflectance data (i.e. low error from atmospheric correction), while the additive nature of IOPs may produce ambiguous results for the inversion algorithms since several IOP combinations may produce the same R_{rs} spectrum (Defoin-Platel and Chami, 2007). As yet there is no single algorithm that can successfully retrieve [Chl a] for all global water types, and given the considerable natural variability of IOPs and their combined ranges, it is unlikely to be attained in future.

1.1.3 South African coastal waters: an ideal case for optical water type classification

With a highly productive coastal upwelling system along the west coast, a shallow shelf environment along the south coast and an oligotrophic western boundary current along the east coast, the coastal waters of South Africa offer a challenging and dynamic bio-optical environment where different water types can occur at varying spatio-temporal scales; these can include optically complex conditions such as high biomass dinoflagellate blooms, coccolithophore blooms, and flood events where standard Case 1 satellite algorithms may fail (e.g. Weeks et al., 2004; Matthews et al., 2012; Smith et al., 2013). The South African coastal environment thus necessitates the use of a variety of algorithms to derive quantitative retrievals of [Chl a] from satellite ocean colour radiometry. OWT classification presents a highly adaptive approach to ocean colour by providing the operational capability for the per-pixel application of the most appropriate chlorophyll algorithms to different optical water types (e.g. Moore et al., 2001; Le et al., 2011; Mélin et al., 2011) whilst improving the overall accuracy of the [Chl a] retrievals for a given region (Moore et al., 2014). The current study will demonstrate the development of such a regional OWT framework (similar to Moore et al., 2014) with the focus on the ability to resolve the dynamic ranges of [Chl a] and regional bio-optical variability found along the coast of South Africa.

1.2 Objectives and key questions

This thesis tests the hypothesis that the satellite-retrieved [Chl *a*] product for South African coastal waters can be improved by the application of a regional OWT classification framework which applies and blends water-type appropriate algorithms per-pixel based on the reflectance spectra. As the only ocean colour sensor with both a ten year archive of historical data and the desired spectral resolution in the red-NIR required for quantitatively resolving high [Chl *a*], the satellite application of the OWT framework focussed on data from the medium resolution imaging spectrometer (MERIS). The following objectives were identified to help test this hypothesis:

- To systematically build an optimized OWT classification framework that is able to characterize the OWTs that may be found in the coastal waters of South Africa.
- To statistically assess the performance of existing Chl *a* algorithms, as well as to assign the best performing algorithm, per OWT.
- To implement these OWTs and Chl *a* algorithms as part of a classification framework for application to MERIS data in order to create a seamlessly blended single optimized [Chl *a*] output product.

Building upon the previously stated objectives, the current body of work will also aim to address the following key questions:

- What is the general persistence of these OWTs in the coastal waters of South Africa over ten years of MERIS data?
- Can the proposed OWT framework correctly identify different water types, and subsequently apply and seamlessly blend the appropriate Chl *a* algorithms at the event-scale in different coastal regions?
- How does the blended [Chl *a*] product compare to currently available regional and standard Chl *a* MERIS products within the different coastal regions of South Africa?
- What are the limitations of the OWT classification and blending framework when applied to MERIS data?

1.3 Thesis structure

Chapter 2 details the data and methodology used in building and determining a representative clustering database aimed at fulfilling specific classification applications. This chapter also presents the various techniques and statistics that were utilized during clustering, classification, and algorithm weighting and blending. The statistics and Chl *a* algorithms that were used to select the most appropriate algorithm per class are also detailed.

Chapter 3 presents the clusters resulting from each of the steps in the systematic clustering database expansion and modification process. A quantitative analysis of the bio-optical causality behind the imposed cluster structures is provided for each step. The clusters from each step of the process are used to classify a selection of regionally-representative test images, whilst the most appropriate set of clusters are determined based on ecologically-focussed [Chl *a*] product applications. The performance of a range of well-known chlorophyll algorithms is statistically assessed per cluster. The best performing algorithm per cluster is utilized in the classification and blending procedure, which is applied to a selection of MERIS match-up data to determine sensor-specific errors and uncertainties.

Chapter 4 demonstrates the application of the final OWTs to the entire MERIS archive for South African coastal waters. The shelf and regional-scale dominance and persistence of individual OWT are discussed in terms of satellite coverage and data quality. Examples of regional event-scale classification with detailed algorithm application and blending are examined. Time series and climatologies of the blended [Chl *a*] product are presented for several coastal, shelf and offshore regions. The performance of the blended product is compared to standard and regional satellite products and discussed in terms of ability to capture the ranges introduced by seasonal and inter-annual variability in surface phytoplankton biomass. This chapter provides an updated analysis of satellite-derived Chl *a* variability for South African coastal waters from an optimized [Chl *a*] product to establish an ecological baseline for the region. It also provides the first comprehensive analysis of key factors affecting the seasonality and regionality of satellite data coverage and quality for this region of the ocean.

Chapter 5 summarizes the conclusions of this thesis and how the outcomes can be used for greatest implication and impact, and provides recommendations for further work that would greatly benefit the utility of OWT classification in South African coastal waters.

Chapter 2

Data description, techniques and methodology

2.1 Introduction

2.1.1 The history of data clustering and classification

The terms *clustering* and *classification* are often used interchangeably throughout the literature, particularly where only one of these methods is used; this may cause some confusion, since they can involve statistically distinct techniques. The following section sets out to explain these terms in the context that they are referred to in this thesis, where clustering (unsupervised classification) is the grouping of unlabelled patterns (i.e. R_{rs} spectra) into meaningful groups (optical water types), whilst classification (supervised classification) uses a previously created collection of labelled patterns to label/classify newly encountered unlabelled patterns (Jain et al., 1999), i.e. satellite R_{rs} spectra. Although there are many different types of clustering and classification algorithms, this section will primarily focus on the types that have been applied to *in situ* bio-optical and satellite ocean colour data.

Data clustering

Clustering is a method of partitioning objects so that the objects in the group are more similar to each other than objects in other groups; the resulting groups are known as clusters. This is a popular and useful step in exploratory data analysis. There are a variety of algorithms to achieve this task and they often aim to minimize the distance between cluster members. The appropriate algorithm, parameter settings and number of clusters depend on the nature of the input dataset as well as the intended use of the

results. Clustering can be distinguished as hard or soft (fuzzy) clustering; with the hard clustering approach every object in the dataset either belongs to a cluster or not, whilst with fuzzy clustering an object can belong to more than one cluster and has a numerical likelihood of belonging (membership) to each cluster. Fuzzy clustering can be converted to hard clustering by assigning each object to the cluster for which it has the highest degree of belonging.

Two of the main clustering approaches include hierarchical and non-hierarchical (partitional) clustering, where hierarchical approaches produce a nested series of partitions and partitional approaches produce only one structure (Jain et al., 1999). Hierarchical clustering algorithms can be either agglomerative or divisive and yield a dendrogram which represents the nested groupings. In agglomerative clustering, for example, at the lowest level each cluster contains one data point, whilst at subsequent levels the most similar clusters are merged; the way in which these clusters are merged (i.e. the characterization of similarity between clusters) may involve various different techniques (e.g single-linkage; complete-linkage), which will not be discussed here. The dendrogram can be broken at different levels to yield different versions of clustering of a dataset. Hierarchical clustering methods have been applied to identify different optical water types from satellite reflectance (e.g. Chen et al., 2004) and radiance (e.g. Yacoub et al., 2001) spectra, *in situ* reflectance spectra (e.g. Lubac and Loisel, 2007; Vantrepotte et al., 2012; Shi et al., 2013; Bao et al., 2015), in addition to being used to discriminate different phytoplankton assemblages from *in situ* reflectance and pigment data (Torrecilla et al., 2011), and to illustrate similarities between different optical water type classes (Mélin and Vantrepotte, 2015).

Partitional clustering algorithms are considered more appropriate for larger datasets where the construction of a dendrogram can be computationally intensive (Jain et al., 1999). A general description of this method is to maximize the cohesiveness within each cluster, while maximizing the heterogeneity among clusters. The user is required to define the amount of output clusters manually; thus the algorithm is often run several times with different starting conditions until the optimal configuration is found (Jain et al., 1999). Several validity functions are available that provide an indication of the optimal number of clusters. Some of the most popular non-hierarchical clustering techniques include k-means and fuzzy c-means (FCM) cluster analysis. K-means has been used to cluster satellite chlorophyll climatology and annual phenology data for biogeographical characterization of the Mediterranean (d'Ortenzio and Ribera d'Alcalà, 2009; Mayot et al., 2015), as well

as to define different optical water types based on *in situ* reflectance (e.g. Feng et al., 2005; Matsuoka et al., 2013) and other optical properties (e.g. Reinart et al., 2003); it has also been applied directly to satellite reflectance spectra as an exploratory diagnostic tool (Karabashev et al., 2006; Karabashev and Evdoshenko, 2016). FCM has most commonly been applied to *in situ* (e.g. Moore et al., 2001, 2014; Grant et al., 2015; Shen et al., 2015) as well as satellite (e.g. González Vilas et al., 2011; Moore et al., 2012) reflectance spectra for the purposes of distinguishing optical water types within the data. Ressom et al. (2006) and Cococcioni et al. (2004) have applied the FCM clustering approach to synthetic data for bio-optical modelling and algorithm development. Clustering is often applied to *in situ* or satellite training datasets for the purposes of one or a combination of the following: parametrizing water type-specific algorithms (e.g. Moore et al., 2001; Feng et al., 2005; González Vilas et al., 2011; Bao et al., 2015), determining water type-specific uncertainty estimates (e.g. Moore et al., 2009; Grant et al., 2015), and to pre-define optical water types for the purpose of satellite image classification and/or class persistence studies (e.g. Vantrepotte et al., 2012; Moore et al., 2014; Mélin and Vantrepotte, 2015).

Ocean classification

Classification is considered to be the ordering of objects based on their similarities to each other or to a set of pre-defined criteria. The classification of the ocean into ecological provinces (Longhurst, 1998) provides a framework for better understanding the governing processes of the various regions and the interactions between them (IOCCG, 2009a). These different regions may be delineated by similarities in physical forcing (e.g. Sathyendranath et al., 1995). In the case of the current study the focus is on the identification of waters with similar bio-optical properties, giving rise to the concept of *optical water types*. The origins of optical classification may be traced back as far as 1865 when Father Pietro Angelo Secchi first used the Secchi disc to measure the transparency of water bodies (Arnone et al., 2004). However, the concept of the quantitative optical water mass classification of the ocean was introduced by Jerlov (1951) and was further refined by Jerlov (1976); these studies classified ocean waters based on the percentage transmittance of downward irradiance, whilst subsequent studies instead used the vertical attenuation coefficient for irradiance (Pelevin and Rutkovskaya, 1977; Smith and Baker, 1978) or absorption spectra (Kirk, 1980; Prieur and Sathyendranath, 1981; Mueller and Lange, 1989; Gould and Arnone, 2003). A popular optical classification theory was proposed by Morel and Prieur (1977) and further developed by Gordon and Morel (1983) and Morel (1988), who classified waters simply as Case 1 and Case 2; Case 1 waters include those where phytoplankton and their derivative products determine the optical properties of the ocean,

whilst Case 2 water optical properties are influenced by other constituents (e.g. suspended inorganic particles and CDOM) whose concentrations do not co-vary with phytoplankton. Although the concept of Case 1 and 2 waters stimulated the development of many of today's standard satellite chlorophyll algorithms, discussions by Mobley et al. (2004) have since motivated the move away from such a simplistic bipartite approach.

The field with the most relevance to the current study is remotely sensed image classification, which has been used for applications such as landcover mapping (Smits et al., 1999; Xie et al., 2008) as well as environmental and water resource monitoring and management (Bastiaanssen et al., 1998; Govender et al., 2007; Giardino et al., 2010). Marine applications have included the retrieval of bottom types and bathymetry in coastal environments (Dierssen et al., 2003; Louchard et al., 2003; Mobley et al., 2005; Benfield et al., 2007), oil spill detection and classification (Solberg et al., 1999; Fiscella et al., 2000), and coastal vulnerability assessment (Dwarakish and Nithyapriya, 2016), among others.

Feature-based classification is a popular method of identifying and extracting relevant characteristics of interest from a predefined feature set (Traykovski et al., 2003). Phytoplankton bloom identification methods frequently make use of feature-based approaches, since there is potential to distinguish different functional types from their unique spectral characteristics (IOCCG, 2014); this characterization is often done via sets of thresholds and/or slopes between specific wavelengths (e.g. Subramaniam et al., 2001; Alvain et al., 2005; McKinna et al., 2011; Siswanto et al., 2013; Dwivedi et al., 2015). Otero and Siegel (2004) developed threshold indices to denote the presence of sediment plumes and non-specific phytoplankton blooms from SeaWiFS water-leaving radiance at 555 nm and [Chl *a*]. Li et al. (2012) and Shi et al. (2013) used boolean criteria in combination with ratios of reflectance measurements at specific wavelengths to distinguish different water types from satellite data of Lake Taihu for class-specific chlorophyll algorithm application. Some studies of coccolithophore blooms have used a parallelepiped algorithm to classify CZCS (Brown and Yoder, 1994; Merico, 2003), SeaWiFS (Iglesias-Rodríguez et al., 2002; Iida et al., 2002) and MODIS (Iida et al., 2012) pixels as *bloom* or *non-bloom* using a set of predefined spectral feature characters. Other classification techniques have used Self-Organizing Feature Maps (SOMs) (Kohonen, 1984), an unsupervised neural network based on competitive learning, which have been applied to classify radiance data over the ocean to distinguish different water and aerosol types (e.g. Ainsworth and Jones, 1999; Yacoub et al., 2001; Niang, 2003). The supervised maximum likelihood classification procedure (Richards, 1993) has been used to distinguish river plumes from other shelf waters

in satellite images (Thomas and Weatherbee, 2006; Lihan et al., 2008), whilst Bao et al. (2015) used the normalized mutual information to identify, classify and apply weighted class-based chlorophyll algorithms in Taihu Lake.

Many methods of classification involve the use of a distance metric to indicate the similarity between two points or vectors, one of which would generally have been predefined during clustering of a training dataset; the new vector can then either be classified into a known class by means of a threshold (i.e. if the distance metric is less than a certain value then the point/vector belongs to the class) or by fuzzy membership (i.e. by quantifying the degree of membership of the new vector to the predefined class). Mayot et al. (2015) used the Chebyshev distance to relate their annual satellite trophic regimes of the Mediterranean to the original bioregions defined by d'Ortenzio and Ribera d'Alcalà (2009). Traykovski et al. (2003) used Euclidean distance and Eigenvector classifiers on remotely sensed water-leaving radiance data of three spectral bands to classify optical water types of the Northwest Atlantic. Many studies have used the Mahalanobis distance to classify ocean colour satellite data (Fiscella et al., 2000; Moore et al., 2001; Alimonte et al., 2003; Moore et al., 2009; Mélin et al., 2011; Vantrepotte et al., 2012; Moore et al., 2012, 2014; Mélin and Vantrepotte, 2015).

The techniques mentioned above are generally only qualitative, but can however provide the framework for the application of water type appropriate algorithms to facilitate accurate quantitative retrievals of the optically significant constituents. Many of the above-mentioned techniques are also hard classification techniques, since they define specific boundaries between different water types that are optically different from each other. The ocean is however a fluid and dynamic environment where such hard boundaries are seldom found; thus fuzzy boundaries may be more applicable. Moore et al. (2001) implemented a probability density function for the Mahalanobis distance to produce fuzzy memberships, or a degree of belonging to a specified water type; the membership functions of satellite pixels can in turn be utilized to blend different algorithm products, in addition to determining per-pixel uncertainty estimates (Moore et al., 2009; Grant et al., 2015; Moore et al., 2015). Many of the methods employed in this thesis are based on the techniques used by Moore et al. (2001, 2014).

2.2 Chapter aims

The aim of this chapter is to present the methods and techniques involved in creating a regionally optimized optical water type classification system for South African coastal waters for application to satellite data, with the purpose of the selection and blending of water type-appropriate algorithms. The first section provides a description and methods of obtaining the *in situ*, synthetic and satellite data used in the composition of the regional clustering database. The second section describes the techniques used for building an appropriate regional dataset, clustering said dataset, and classification of satellite imagery. The third section outlines the assessment of the dominant modes of variance and optical causality of the various data sources on clustering and classification performance. The last section describes the chlorophyll algorithms, error and uncertainty assessment techniques of the various algorithm outputs when applied to *in situ* and satellite reflectance data, and the final algorithm selection and blending procedure.

2.3 Data and methods

2.3.1 Data description

The successful optical characterization of a study region requires thorough consideration of the spatial and temporal variability that may be encountered; incorporating the different levels of expected optical variability in the training dataset may increase the representativeness of the resulting clusters. The following section details the data utilized in the clustering database in the current study.

In situ data

The majority ($\approx$ 78 %) of the *in situ* data were collected in the southern Benguela along the west coast of South Africa; this region has been the focus of many ocean colour remote sensing studies since 2002 due to the high productivity of the upwelling system and resulting harmful algal blooms. Field campaigns have most often focussed on the upwelling or high productivity seasons (February to April) in order to capture *in situ* and satellite validation data for the phytoplankton blooms which often occur during this time; as a result the *in situ* data show a strong seasonal and regional bias with a paucity of data available for the south and east coast of the country. The available data were collected during collaborative research efforts between the Department of Agriculture, Forestry and Fisheries (DAFF), the Council for Scientific and Industrial Research (CSIR) and the University of Cape Town (UCT) and have included data collection in the St Helena Bay region near Lambert's Bay (N=147) and Elands Bay (N=37), in Saldanha Bay (N=10) during 2012 and 2013, in Algoa Bay during May 2012 (N=5), and in the Natal Bight during November 2010 and May 2011 (N=54). The methodological details for these field studies were all similar, and are described below. Additional data collected in the Benguela region includes the Benguela Calibration (BENCAL) cruise (N=20); details of the data collection methodology can be found in the cruise report (Barlow et al., 2003). An *in situ* dataset which focussed on the R_{rs} and [Chl *a*] data were compiled from the aforementioned studies.

Radiometric data

In-water radiometric measurements were made with a Satlantic Hyperspectral Teathered Radiometric Buoy (TSRB). The TSRB measures upwelling radiance ($L_u(z)$ at z = -0.66 m, μW cm^{-2} nm^{-1} sr^{-1}) and above surface downwelling irradiance ($E_d(0^+)$, μW cm^{-2} nm^{-1}) and has two 256 channel spectrographs that are linked to a downward

facing 8.5° field of view radiance sensor and an upward looking cosine corrected irradiance sensor. The two spectrographs covers a spectral range of 400 to 800 nm, with a sampling distance of 3.3 nm to an accuracy of 0.3 nm. Acquisition rates vary between 0.7 and 1.6 Hz in response to the light field. During aquisition the instrument was floated far enough from the vessel to avoid shadowing or interference. Measurements were typically recorded for about 2 to 5 minutes. Raw data were processed with Prosoft 6.3d (Satlantic: Halifax, Canada); the median values of the deployment were calculated and resampled to a spectral resolution of 5 nm.

In the absence of systematically measured *in situ* IOP data, the Equivalent Algal Population (EAP) inversion algorithm using Ecolight-S inversion (Evers-King et al., 2014) was applied to the *in situ* radiometric data in order to obtain estimates for the IOPs. Inputs to this model are hyperspectral, subsurface radiance (L_u(-0.66 m)(λ), μW cm^{-2} nm^{-1} sr^{-1}) and hyperspectral, above surface downwelling irradiance ($E_d(0^+)(\lambda)$, μW cm^{-2} nm^{-1}). The EAP inversion algorithm considers four major components: water, phytoplankton, non-algal particles, and combined CDOM and detritus; in the model water and phytoplankton contribute to both absorption and backscattering, whilst non-algal particles and combined CDOM and detritus contribute only to backscattering and absorption respectively (Evers-King et al., 2014). The various models that were used for each of the components are discussed further in Evers-King et al. (2014). The EAP algorithm has been validated and tested against the phytoplankton inherent optical property (IOP) models from IOCCG (2006) and Alvain et al. (2012) and was found to successfully simulate reflectance across a range of biomass and phytoplankton assemblage characteristics (Robertson Lain et al., 2014).

Optional outputs to the Ecolight-S inversion include spectra of R_{rs} as well as several IOPs, which were assimilated into the clustering database; these IOPs included:

- combined CDOM and detrital absorption a_{dg} (m^{-1})
- phytoplankton absorption a_ϕ (m^{-1})
- phytoplankton backscatter $b_{b\phi}$ (m^{-1})
- non-algal particulate backscatter b_{bs} (m^{-1})

Pigment analysis

Chlorophyll *a* concentration ([Chl *a*]) was measured by fluorometric analysis with the use of a Turner Designs 10-AU Fluorometer according to the protocols described by Mueller et al. (2003), which in turn are based on the JGOFS Protocols (Knap et al., 1996). Discrete water samples of volumes between 0.1 and 1 litre were filtered through 25 mm Whatmann GF/F glass fibre filters subject to 10 mm mercury pressure. Filtered sample papers were folded and placed in polypropylene tubes with 9 ml acetone (90%), ground with a glass rod for one minute and then frozen for 24 hours to allow for pigment extraction. The test tubes were then centrifuged at 2500 rpm for 10 minutes to reduce turbidity, after which the supernatant was transferred to 13 mm x 100 mm disposable culture glass tubes to be read in the fluorometer. The fluorometer was zeroed using 90% acetone prior to taking readings. The supernatant was first read in the fluorometer, and then acidified to correct for phaeophyton pigments with 0.15 ml 0.2N HCL solution using a Pasteur pipette, and allowed to stand for one minute before being re-read in the fluorometer. The corrected [Chl *a*] was calculated using the following formulae:

$$[\text{Chl } a](\text{mg m}^{-3}) = a_c * (R_b - R_a) * \frac{\text{extraction volume}}{\text{sample volume}} * DF \tag{2.1}$$

where a_c is the calibration coefficient for the fluorometer, R_b is the fluorometer reading before acidification, R_a is the fluorometer reading after acidification, and DF is the Dilution Factor (=1).

Synthetic data

The existing *in situ* data were augmented with hyperspectral synthetic data as a first attempt to increase the dynamic range of the clustering database. Synthetic data are useful since the created R_{rs} spectra can be deconstructed into IOPs from well understood and validated models. Synthetic data are also not subject to measurement mismatch or processing errors, which can be a common occurrence in *in situ* data collections. Relevant and useful synthetic spectra should represent water constituents and IOPs, in addition to ranges of covariation between IOPs, that occur naturally in the ocean or a given region (IOCCG, 2006; Defoin-Platel and Chami, 2007). It should be noted that synthetic data will not necessarily correspond to all natural situations; however, adding constraints to account for the appropriate ranges of variation of the IOPs as well as the possible covariation between different IOPs, will aid the generation of more realistic synthetic data (Defoin-Platel and Chami, 2007). The models and parameters used to derive the IOPs

and AOPs are generally derived from field measurements, meaning that the resulting data should be consistent with a wide range of naturally occurring waters.

The synthetic data used in this study were assimilated from two sources: the IOCCG Report 5 (IOCCG, 2006), and a new Case 2 dataset created specifically for this study with Ecolight. The same set of IOPs were extracted for the synthetic data as for the *in situ* radiometric data. The full dataset that is used to create the OWTs comprises of approximately 39 % synthetic data. Further details of the respective models used to create the synthetic data are provided in appendix A.

Extracted satellite data

Satellite data were obtained from the Medium Resolution Imaging Spectrometer (MERIS) onboard the polar orbiting Envisat Earth Observation Satellite from the European Space Agency (ESA). MERIS is a passive optical pushbroom wide-field instrument that has five identical cameras which image across-track stripes of the earth's surface. It covers a swath of 1150 km and has a revisit time of 2-3 days. The reduced resolution (RR) data has a spatial resolution of 1040 x 1200 m. Optical data are collected in the 390 - 1040 nm range of the electromagnetic spectrum at 15 spectral bands with average bandwidths of 10 nm; only the first ten bands, centered at 412.5, 442.5, 490, 510, 560, 620, 665, 681.25, 708.75 and 753.75 nm, were of interest to this study.

Level 2 RR data from the 3rd reprocessing were obtained from the MERIS catalogue and inventory (MERCI) for the period ranging from April 2002 to April 2012. This included water-leaving reflectance ρ_w for the first 10 bands, data quality flags, as well as the Algal Pigment Index I (Algal1) and II (Algal2) which represents the [Chl a] products for Case 1 and Case 2 waters respectively. The ρ_w was converted to R_{rs} in order to be compatible with *in situ* radiometric data and clusters as follows (Antoine and Morel, 2005):

$$R_{rs} = \frac{\rho_w}{\pi} \tag{2.2}$$

A variant of the *in situ* EAP algorithm from Evers-King et al. (2014) was applied to the Level 2 MERIS data as described in Bernard et al. (2014). The algorithm is essentially the same as the one applied to the *in situ* R_{rs}, without the depth dependency; satellite R_{rs} is used as the input parameter for the minimisation procedure. The inversion requires a set of initial starting conditions; these included using the Algal1 data for [Chl a], effective diameter of 10 μm, solar zenith angles from the Level 2 data file, while a_{gd} and b_{bs} were

calculated proportional to the empirically estimated [Chl a]. Where applicable, negative R_{rs} were replaced with near zero values, while an empirical relationship between 665 and 709 nm (Bernard et al., 2005) was used for conditions where Algal1 values suggested [Chl a] >10 mg m^{-3}.

Where satellite data were extracted for addition to the clustering database, the satellite R_{rs} spectra and [Chl a], together with the following associated EAP model outputs, were used: $a_{dg}(440)$, $b_{bs}(560)$, $a_{\phi}(440)$, $a_{\phi}(680)$, $b_{b\phi}(560)$ and $b_{b\phi}(710)$. The full dataset that is used to create the OWTs contains approximately 47 % extracted satellite data. The extracted [Chl a] data that were utilized depended on the estimations of the Algal1 product, and whether the specific reflectance spectrum fell in the moderate or elevated reflectance groups (these terms and their definitions are discussed later in this chapter in section 2.3.2): for the moderate reflectance spectra Algal1 was used for any estimates of <10 mg m^{-3}, whilst the EAP output were used for pixels with Algal1 estimates of concentrations >10 mg m^{-3} (Evers-King, 2014); for the elevated reflectance spectra the Algal2 product was used.

2.3.2 Data clustering and classification

The data clustering and classification procedure takes the form of a stepwise analysis, where the input dataset is systematically expanded and modified until satisfactory clusters (with regard to the desired classification applications) are achieved. A flowchart outlining the procedure is given in figure 2.1.

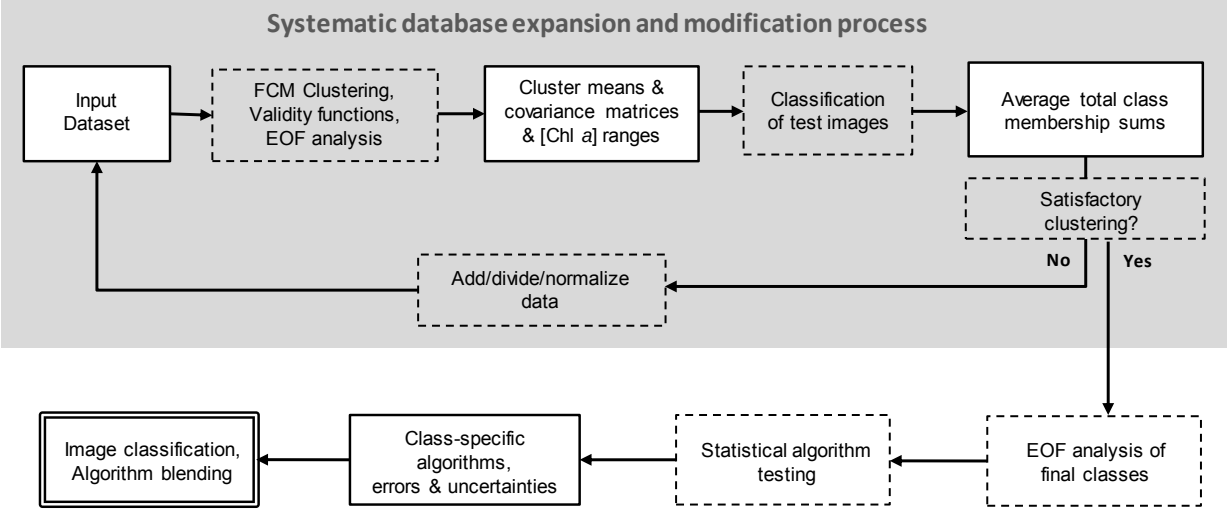


Figure 2.1: Flow chart outlining the clustering and classification methods. The solid line boxes represent specific inputs or outputs. The dashed boxes represent the techniques that were applied to the data.

Data clustering

An unsupervised fuzzy c-means (FCM) cluster algorithm (Bezdek, 1981) was applied to a selection of seven R_{rs} datasets. In this case the FCM algorithm from MATLAB (Math Works Inc.) was used. The algorithm minimizes an objective function J_m defined as:

$$J_m = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m \|x_k - v_i\|^2 \tag{2.3}$$

where u_{ik}, or $u_i(x_k)$, is the membership of the kth observation to the ith cluster, $\|x_k - v_i\|$ is the Euclidean norm between vectors x_k and v_i, m is the weighting exponent that can be any real number greater than 1, c is the number of clusters and N is the number of observations. Within this thesis the application of the algorithm minimizes the distance between data points x_k (a R_{rs} spectrum) and the prototype cluster centers v_i, iteratively adjusting the cluster centers until optimization criteria are met. Apart from m the default settings were used for the FCM function, which included a maximum of 100 iterations or a minimum objective function improvement of 10^{-5} between consecutive iterations.

Pal and Bezdek (1995) recommend caution when using m of less than 1.5 or greater than 2.5; however Moore et al. (2001) found that 1.2 and 1.5 indicated the same amount of clusters for their dataset, and ended up using 1.2. In the current study the use of m >1.2 tended to cause increased cluster overlap; thus 1.2 was chosen for this study.

Since the optimal value for c is not known prior to clustering, a selection of four validity functions were calculated for a range of c between 2 and 20 in order to determine the appropriate amount of clusters for the dataset. These included the Partition Coefficient, Classification Entropy, Partition Index and Separation Index as detailed below. Validity functions provide an evaluation of the performance of the clustering by means of a quantitative measure of the quality of the outputs (Windham, 1982) and the significance of the structure imposed on the data (Xie and Beni, 1991).

The Partition Coefficient (PC) (Bezdek, 1974) is an indicator of the separation of clusters and measures the overlap between clusters. The closer this value is to one the better the data were classified (Windham, 1982). PC does not consider the data structure, is sensitive to the values of c and m (Windham, 1982) and thus has a tendency to decrease with increasing c (Xie and Beni, 1991). PC is defined as:

$$PC = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}(u_{ij})^2 \tag{2.4}$$

where u_{ij} is the membership of the jth observation to the ith cluster.

The Classification Entropy (CE) (Bezdek, 1974, 1975) measures the fuzziness of the cluster partition; a desirable outcome is close to zero. Similar to PC, the CE suffers from sensitivity to its input parameters (Windham, 1982) and has no direct connection to the geometric properties of the data (Xie and Beni, 1991). It is defined as:

$$CE = -\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}u_{ij}log(u_{ij}) \tag{2.5}$$

The Partition Index (SC) (Bensaid et al., 1996) is a measure of the compactness and separation of the clusters. A lower SC indicates a better partition, and it is defined as:

$$SC = \sum_{i=1}^{c}\frac{\sum_{j=1}^{N}u_{ij}^m\|x_j - v_i\|^2}{N_i\sum_{k=1}^{c}\|v_i - v_k\|^2} \tag{2.6}$$

where N_i is equal to $\sum_{j=1}^{N} u_{ij}$ which is the fuzzy cardinality of each cluster. Normalization by N_i is performed to make SC insensitive to cluster sizes (Bensaid et al., 1996).

The Separation Index S (Xie and Beni, 1991), also known as the Xie-Beni Index, aims to quantify the ratio of the total variation within clusters and the minimum-distance separation of cluster centers; it favours clusters that are maximally separate from one another (Bensaid et al., 1996). S is directly related to the geometric properties of the clustering dataset and proportional to the overall average compactness and separation of clusters (Xie and Beni, 1991). S is defined as:

$$S = \frac{\sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^{m} \|x_j - v_i\|^2}{N * min_{i,k} \|v_i - v_k\|^2} \tag{2.7}$$

The results from all the validity functions for each of the c from 2 to 20 were plotted, and taken into consideration when choosing the appropriate amount of clusters for each dataset. R_{rs} spectra were grouped into the class to which it had the highest membership; all data related to a specific R_{rs} spectrum (e.g [Chl a], a_{dg}, a_{ϕ}, $b_{b\phi}$, b_{bs}) were also sorted into the respective clusters to be used for further analysis.

Satellite image classification

The image classification techniques are based on the concept of fuzzy logic (Zadeh, 1965), and uses the methods of Moore et al. (2001, 2009, 2012, 2014). In terms of satellite imagery, the basic concept is that the observed reflectance spectrum from a pixel is given an index of similarity or membership to each cluster. The squared Mahalanobis distance Z_i^2 was used to quantify the closeness between the observed spectra and the cluster mean:

$$Z_i^2 = (R - M_i)^t \Sigma_i^{-1} (R - M_i) \tag{2.8}$$

where R is the observed spectrum, M_i is the mean vector of the ith class, Σ_i is the covariance matrix of the ith class, and t indicates the transpose of the vector $(R - M_i)$.

The measure of the likelihood that a spectrum is in class i is given by the membership function f_i :

$$f_i = 1 - F_n(Z_i^2) \tag{2.9}$$

where $F_n(Z^2)$ is the cumulative χ^2 distribution function with n degrees of freedom (in this case the 9 MERIS spectral bands). The output f_i is a value between 0 and 1 which

describes the level of similarity between an observation and the mean vector of class i; if the observed spectrum is identical to a class mean, it will have membership value of 1. The membership function is otherwise not constrained and does not have to sum to one.

The classification of MERIS images only used wavebands 2 to 10. The 412 nm band was not used during classification, as done in previous studies (Vantrepotte et al., 2012; Moore et al., 2014), since this band in MERIS data from the 3rd reprocessing is generally noisy (Smith et al., 2013; Zibordi et al., 2013; Cristina et al., 2014).

Determination of clustering database and final classes

Seven different input datasets were used during the clustering and classification scheme outlined in figure 2.1. The methodology involved a systematic process wherein the initial clustering was performed on the *in situ* data, whilst each subsequent step was augmented with synthetic and/or extracted satellite data, or modified by division or normalization. In each step the relative clustering success was determined and compared to previous steps to establish the optimal clustering dataset and number of clusters for satellite image classification. These datasets are described in figure 2.2.

The validity functions shown in equations 2.4 to 2.7 were used to detect the optimal number of clusters for each dataset. Each dataset underwent FCM cluster analysis in order to calculate class statistics (mean and covariance matrix for each cluster), which were in turn used to classify a group of test images for three domains as represented in figure 2.3: west coast, south coast and east coast. These test images were spread over 10 years and included data from different months in an attempt to account for possible seasonal in-water and atmospheric changes which could have an impact on the classification. Ten images were selected from each domain (the dates and file names for each of the domains can be seen in tables C.1, C.2 and C.3). The average total class membership sum (A) was calculated for each image; A is defined as:

$$A = \frac{\sum_{i=1}^{c} \sum_{p=1}^{n} f_i p}{\sum_{p=1}^{n} p} \tag{2.10}$$

where n is the number of valid pixels p in each image, whilst c is the number of clusters and $f_i p$ is the membership function of the n^{th} valid pixel to the i^{th} cluster. An A value of close to or above 1 is desirable since it means that all or most spectra from the image were represented by one or more of the clusters of the particular dataset. The mean ($\bar{A}$) and standard deviation of A from all ten images were noted for each dataset and domain (table

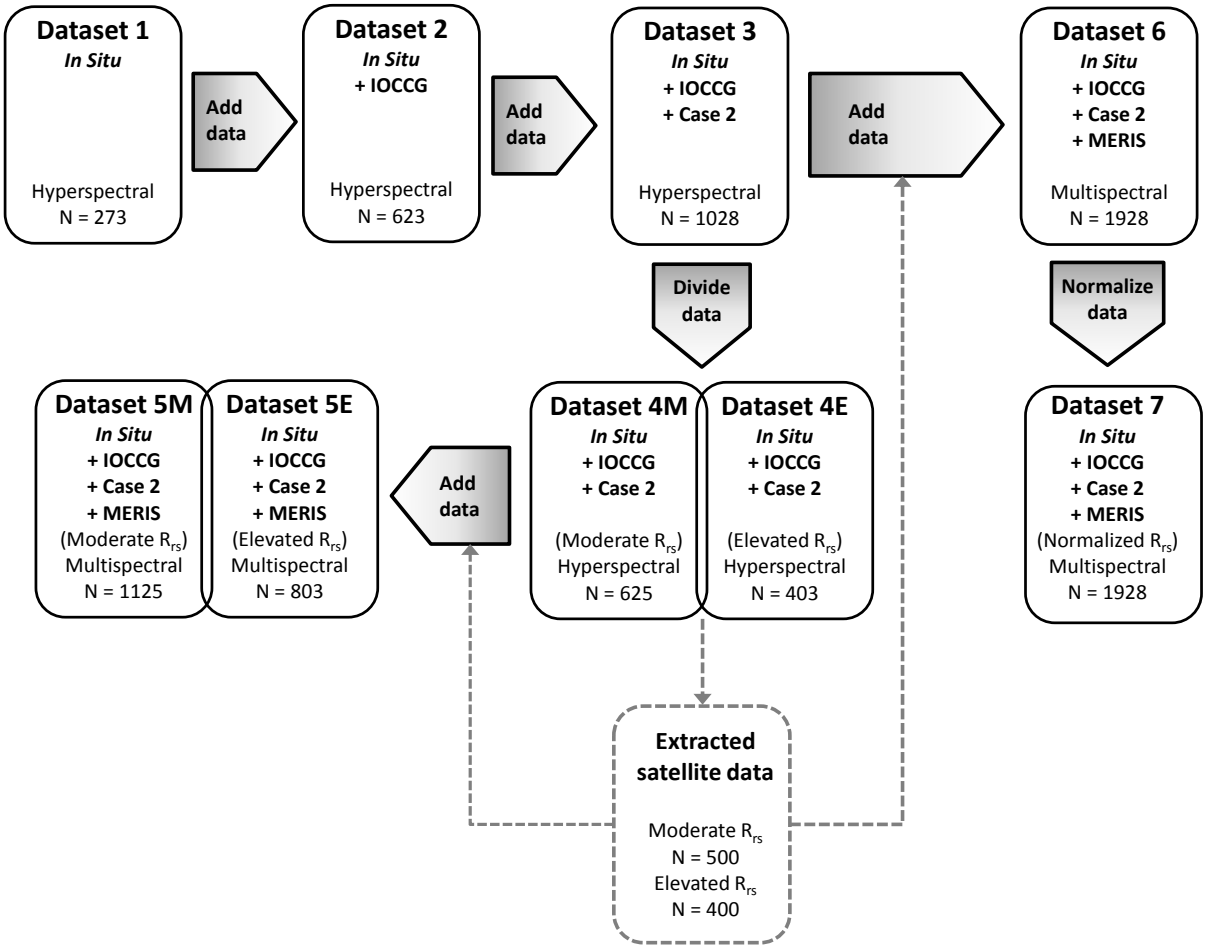


Figure 2.2: Diagram describing each of the clustering datasets. Each block outlines the source data, spectral resolution, and the total number of spectra in each dataset. The grey arrows indicate the processes used to create subsequent datasets.

3.1 in the results); thus $\bar{A}$ was used as a measure of the overall classification success of each dataset and cluster grouping per domain. A similar measure of classification success was used in Mélin and Vantrepotte (2015), although they calculated their domain average of the total class membership using all available classified satellite data over the seven year study period. Several factors could affect the classification success of an image, which will be discussed later in chapters 3 and 4; however, applying iterations of different cluster sets over the same set of test images could give an indication of the improvement in the representativeness of the clustering database and the data groupings. High membership sums, and thus a high A value, indicates that one or many of the classes are represented in the database, whilst low membership indicates that the clusters are not parametrized correctly, or do not represent the optical water types of the region.

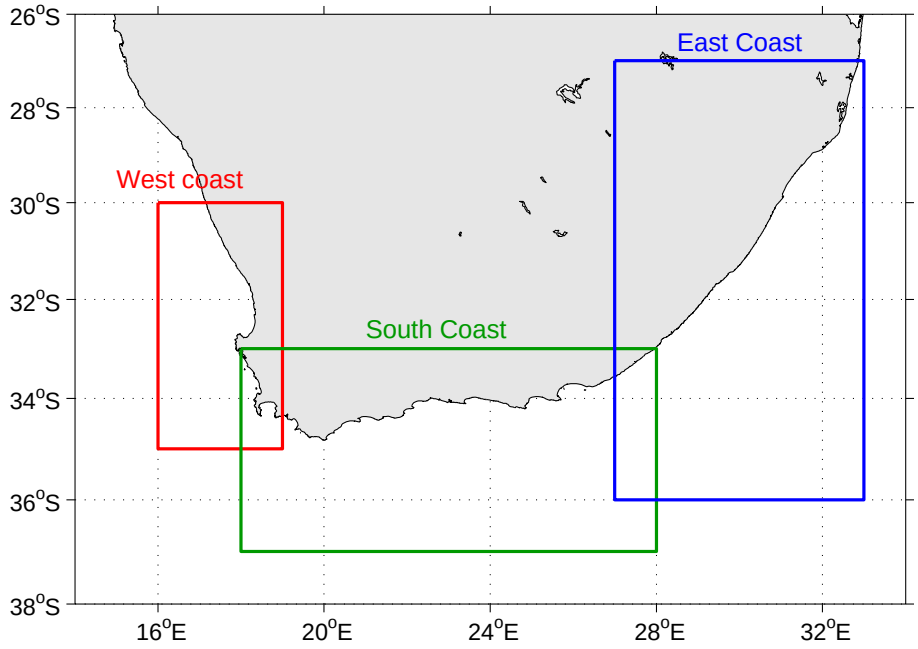


Figure 2.3: A map showing the bounding boxes for the three regions of interest, including the west coast (red), south coast (green) and the east coast (blue).

After iteration 3 the clustering database was separated into a moderate and an elevated R_{rs} dataset. This division was based on reflectance ratios in the blue and green spectral regions; spectra that showed either $R_{rs}(560) > 0.009$, or $R_{rs}(490) > 0.009$ together with $R_{rs}(490)/R_{rs}(443) > 1$, were added to the elevated reflectance database, whilst the rest were added to the moderate reflectance database. This division was performed in order to avoid biasing the cluster partitions towards the relatively large spectral changes in shape and magnitude that occur at elevated R_{rs}; the reasoning and methods behind the division are discussed further in section 3.3.2. The two resultant clustering databases were subsequently clustered individually in iteration number 4 and 5.

MERIS Level 2 data were extracted and added to the clustering database after iteration 4. Since EAP data and associated IOPs were only available for the Southern Benguela, only data from this region were extracted. Data were extracted from a selection of 24 images covering 8 years worth of data and different months (the dates and file names for each of these files can be seen in table C.4). The MERIS reflectance quality flag, PCD1_13, was applied so that only pixels with valid reflectance could be used, whilst only data from areas with membership sums of less than 0.1 were extracted; these criteria were set in an attempt to improve classification success by finding and adding missing spectral shapes to the clustering database. These extracted data were divided according

to the above-mentioned reflectance thresholds and ratios into the elevated and moderate databases. For each of the extracted data the associated EAP IOPs were also retrieved. The MERIS Algal1 product was used to represent satellite [Chl *a*] where the product showed values <10 mg m^{-3}; where Algal1 was flagged or indicated values >10 mg m^{-3} the EAP chlorophyll product was used instead.

There were 477,895 and 13,052 spectra extracted for the moderate and elevated groups respectively. The moderate reflectance spectra were sorted into six groups according to ascending [Chl *a*]; the following concentration ranges were used to define the groups:

- <0.15 mg m^{-3}
- 0.15 - 0.8 mg m^{-3}
- 0.8 - 5 mg m^{-3}
- 5 - 75 mg m^{-3}
- >75 mg m^{-3}

These ranges were roughly based upon the concentration ranges for the biomass clusters resulting from dataset 4, with the addition of an extreme phytoplankton bloom group. A threshold of >75 mg m^{-3} was used to define the phytoplankton bloom spectra; this threshold has previously been used to distinguish phytoplankton blooms in upwelling regions such as California (Ryan et al., 2008) and the southern Benguela (Evers-King, 2014). 100 spectra were selected randomly from each of the five groups and added to the moderate cluster database, whilst 400 MERIS spectra were randomly selected from the elevated group and added to the elevated database; this formed dataset 5.

For iteration 6 the clustering was performed on all the *in situ*, synthetic and extracted satellite data without a database division. Iteration 7 consisted of the same data as iteration 6 but used a larger number of clusters; this was done to demonstrate the focus of the FCM clustering algorithm.

In order to focus on the spectral shape instead of the inherent variability caused by backscattering, dataset 8 comprised a normalized dataset, as used in previous studies (Lubac and Loisel, 2007; Vantrepotte et al., 2012; Shi et al., 2013; Bao et al., 2015; Mélin and Vantrepotte, 2015) following the formula:

$$R_n(\lambda) = \frac{R_{rs}(\lambda)}{\int_{\lambda_2}^{\lambda_1} R_{rs}(\lambda) d\lambda} \tag{2.11}$$

where R_n (in units of nm^{-1}) indicates the normalized spectrum computed by trapezoidal integration between $\lambda_1 = 412$ nm and $\lambda_2 = 753$ nm. Clustering was performed on the R_n spectra. For classification, the class statistics μ and $\sum$ of each resulting cluster were calculated with log-transformed R_n (Alimonte et al., 2003; Mélin et al., 2011; Vantrepotte et al., 2012). Satellite R_{rs} spectra were also normalized to obtain R_n, and subsequently log-transformed. As in previous iterations, the Mahalanobis distance and cumulative χ^2 distribution was used to define class membership (Mélin and Vantrepotte, 2015).

Iterations 1 through 4 of the data clustering and classification scheme were performed on hyperspectral data ranging from 400 to 750 nm with 5 nm increments. Since iteration 5 to 8 contained extracted MERIS data, the hyperspectral data were resampled to MERIS wavelengths prior to clustering. The results from testing the success of each database can be seen in table 3.1 in section 3.2.1.

2.3.3 EOF analysis of cluster datasets and optical water type classes

Empirical orthogonal function (EOF) analysis, also known as principal component analysis (PCA), was used to assess the variance structure within the cluster datasets and the final optical water type classes. The analysis technique reduces the dimensionality of a dataset by breaking it down into a set of geometrically independent (orthogonal) modes of oscillation which represent all the variability in the data (Craig et al., 2012). The first mode represents the largest percentage of the total variance while each subsequent mode accounts for progressively less variance. EOF analysis has previously been used to assess the variance structure of spectral data such as R_{rs} or radiance (Mueller, 1976; Sathyendranath et al., 1989, 1994; Flink et al., 2001; Toole and Siegel, 2001) and a_ϕ (Carver et al., 1994) in addition to being used as a tool to parameterize ocean colour algorithms (Lubac and Loisel, 2007; Fichot et al., 2008; Craig et al., 2012; Bracher et al., 2015).

The PCA function from MATLAB was applied to an input matrix where rows correspond to observations (R_{rs} spectra) and columns correspond to variables (wavelengths). The output included the EOFs or vectors of principal component loadings (coefficients), where each vector contains the loadings for one principal component (mode), with one loading for each wavelength; these modes were organized in descending order of component variance. A vector of scores was also obtained for each EOF, which represents the amplitude of the EOF for each R_{rs} spectrum and defines the directions of maximum sample variance (Lubac and Loisel, 2007). The interpretation of the modes may be complicated, since each mode may be influenced by more than one process or property (Toole and Siegel, 2001). A correlation analysis was performed between the amplitude factors (scores) of the first three modes and a selection of bio-optical parameters ([Chl a], $a_{dg}(440)$, $a_\phi(440)$, $a_\phi(680)$, $b_{b\phi}(560)$, $b_{b\phi}(710)$, $b_{bs}(560)$) in order to determine the IOPs or processes that influence each mode.

2.3.4 Class-specific Chl *a* algorithms

This section describes the assessment and selection process of the appropriate chlorophyll algorithms for each class. As done in Moore et al. (2014), the goal was not to create new algorithms, but rather to utilize existing and validated algorithms and assess their performance per class. The algorithms included three blue-green band ratio algorithms and two red band ratio algorithms.

The first three algorithms utilize the blue and green spectral wavelengths and are band-ratio algorithms in a 4th order polynomial exponential form. The algorithm is defined as:

$$[\text{Chl } a](\text{mg m}^{-3}) = 10^{(A0+A1*X+A2*X^2+A3*X^3+A4*X^4)} \tag{2.12}$$

where $A0$ to $A4$ represent the five different coefficients of the polynomial. For the first two algorithms X is the log10 of the maximum ratio of $\rho_w(443)$, $\rho_w(490)$, $\rho_w(510)$ to $\rho_w(560)$. The first version is the MERIS algal pigment 1 algorithm, OCMe (Morel and Antoine, 2011), whilst the second is the OC4E algorithm (Feldman and McClain, 2015); these two algorithms differ only in the coefficients that they use. The third algorithm, OC3E, uses only three wavelengths; X is the log10 of the maximum ratio of the couple $\rho_w(443)$, $\rho_w(490)$ to $\rho_w(560)$. OC4E is equivalent to the NASA OC4 algorithm (version 6) (O'Reilly et al., 1998) which was designed for SeaWiFS data using R_{rs} at 443, 489, 510 and 555 nm, whilst OC3E is equivalent to the MODIS OC3M (version 6) algorithm which uses the R_{rs} at 443, 489 and 547 nm; both OC4E and OC3E were derived for MERIS wavelengths from version 2 of the NOMAD dataset. The coefficients (shown in table 2.1) from Morel and Antoine (2011) were used for OCMe, whilst coefficients for OC4E and OC3E were retrieved from Feldman and McClain (2015).

Table 2.1: The coefficients for the OCMe, OC4E and OC3E algorithms

	OCMe	OC4E	OC3E
A0	0.450	0.3255	0.2521
A1	-3.259	-2.7677	-2.2146
A2	3.523	2.4409	1.5193
A3	-3.359	-1.1288	-0.7702
A4	0.949	-0.4990	-0.4291

The aforementioned algorithms were designed for open-ocean regions, but it has been suggested that fluorescence-based models or algorithms that use reflectances from the

red and near-infrared (NIR) spectral regions are preferable for turbid productive waters (Gitelson, 1992; Dall'Olmo et al., 2003; Dall'Olmo and Gitelson, 2005) since these bands are less sensitive than the traditional blue-green (440-550 nm) ratio algorithms to the absorption by CDOM and scattering by mineral particles (Gilerson et al., 2010). There also tends to be more signal in the red (compared to the blue-green) at [Chl *a*] above approximately 25 mg m^{-3} (Robertson Lain et al., 2014). Two examples of such models included:

- The two-band model:

$$[\text{Chl } a] \propto (R_{\lambda_1}^{-1}) * R_{\lambda_2} \tag{2.13}$$

- The three-band model:

$$[\text{Chl } a] \propto (R_{\lambda_1}^{-1} - R_{\lambda_2}^{-1}) * R_{\lambda_3} \tag{2.14}$$

When applied to MERIS data, R represents ρ_w, whilst λ_1, λ_2 and λ_3 represents the spectral bands of 665, 708 and 753 nm respectively. Whilst operating in the fluorescence domain, these algorithms do not focus on the height of the fluorescence peak, e.g. the FLH algorithm (Gower et al., 1999), but rather targets the change in the a_ϕ and $b_{b\phi}$ associated with increasing phytoplankton biomass. These algorithms have been regionally calibrated for various study sites (Koponen et al., 2007; Moses et al., 2009a; Matthews et al., 2010). However, the specific calibration parameters applied to data in the present study originated from specialized versions of the 2-band (G2B) and 3-band (G3B) algorithms (Gilerson et al., 2010), which were both calibrated using an extensive synthetically generated dataset. Gilerson et al. (2010) found that these models were sensitive to changes in Chl *a* specific phytoplankton absorption, primarily at [Chl *a*]<10 mg m^{-3}; however, at [Chl *a*]>10 mg m^{-3} these algorithms should not require regional tuning (Gilerson et al., 2010). The algorithms applied in this thesis were:

- The MERIS G2B algorithm:

$$[\text{Chl } a](\text{mg m}^{-3}) = \left[35.75 * (R_{665}^{-1} * R_{\lambda_{708}}) - 19.3\right]^{1.124} \tag{2.15}$$

- The MERIS G3B algorithm:

$$[\text{Chl } a](\text{mg m}^{-3}) = \left[113.36 * \left((R_{665}^{-1} - R_{708}^{-1}) * R_{753}\right) + 16.45\right]^{1.124} \tag{2.16}$$

All blue-green band ratio Chl *a* algorithms were tested on the *in situ* and all (IOCCG and Case 2) synthetic data of each cluster. The algorithms were applied to the hyperspectral data at the approximate MERIS wavebands of 440, 490, 510, 560, 665, 710 and 750 nm; the resultant modelled [Chl *a*] were compared to the available known *in situ* and synthetic [Chl *a*]. The statistical descriptors used to select the appropriate algorithm included the bias (the average difference), median absolute relative error (MARE) and the root-mean-square error (RMSE).

The bias is defined as:

$$\text{bias} = \frac{1}{N}\sum_{i=1}^{N}(\log10Chl_{mod} - \log10Chl_{insitu}) \tag{2.17}$$

the median absolute relative error (MARE) is expressed in percentage as:

$$\text{MARE} = 100 * median\left[\frac{|Chl_{mod} - Chl_{insitu}|}{Chl_{insitu}}\right] \tag{2.18}$$

and the root-mean-square error (RMSE) as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log10Chl_{mod} - \log10Chl_{insitu})^2} \tag{2.19}$$

where Chl_{mod} is the algorithm-derived [Chl *a*] output and Chl_{insitu} is the *in situ* [Chl *a*]. Since the natural distribution of [Chl *a*] in the ocean is considered to be log-normal (Campbell, 1995), the calculation of the bias and RMSE were performed on log-transformed data, as is often done prior to statistical analysis of [Chl *a*] and IOP data (e.g. Cannizzaro and Carder, 2006a; Marrari et al., 2006; Antoine et al., 2008; Moore et al., 2009; Szeto et al., 2011; Craig et al., 2012; Liu and Wang, 2013; Moore et al., 2014). The abovementioned statistics were used to choose the most appropriate Chl *a* algorithm for each cluster; although the desired result would be low values of all three statistics, the best indication was generally given by the RMSE since it is an inclusive metric that combines information of the mean and standard deviation in one term (Szeto et al., 2011). The Chl *a* algorithms were applied to all *in situ* and synthetic data within a cluster and the algorithm with the best statistical performance was assigned to the cluster. Some algorithms were assigned to more than one cluster. The statistics associated with the chosen algorithm were subsequently used to derive the uncertainty associated with each blended [Chl *a*] product.

Algorithm blending

In order to create a blended algorithm product each algorithm requires a weighting factor; the sum total of these weighting factors should amount to one. The weighting factors were created by normalizing the class memberships associated with each algorithm by the total membership according to the following equation:

$$w_i = \frac{\sum_1^x f_n^i}{\sum_1^y f_c} \tag{2.20}$$

where w_i is the normalized weighting factor that will be applied to algortihm i, f_n^i is the fuzzy membership to optical water class n assigned to algorithm i, and the sum in the denominator is over all fuzzy memberships. Note that each reflectance spectrum has a different set of fuzzy memberships, and thus a different set of weighting factors. The blended product is then:

$$Chl_{blend} = \sum_{i=1}^{x} w_i Chl_i \tag{2.21}$$

where Chl_i is the chlorophyll product from algorithm i, and x is the number of different algorithms that are in use. During the blending procedure it is assumed that the satellite reflectance vector has membership to at least one of the defined water types. If this is not the case the pixel is given a null value.

Error and uncertainty characterization of chlorophyll algorithm application to satellite data

In order to characterize the errors and uncertainties inherent to satellite data application, each of the chlorophyll algorithms (including the blending algorithm) were applied to two different satellite validation matchup datasets. These datasets consisted of 48 datapoints collected in the southern Benguela and 241 from the MERis MAtchup Insitu Database (MERMAID) (Barker et al., 2008), respectively. The MERMAID dataset comprised all data where *in situ* [Chl *a*] measurements were taken, which included data from the Black Sea (BioOptEuroFleets), western Mediterranean (BOUSSOLE), Bristol Channel and Irish Sea, Beaufort Sea (CASES), eastern English Channel, French Guiana, NOMAD, California (PlumesAndBlooms), Portuguese coast and French coast (Ifremer-REPHY). Only matchup data that included water-leaving reflectance (or R_{rs}) for MERIS bands 2 to 10 were used, since these are the bands used for spectral classification. The MERMAID dataset is representative of low to medium biomass conditions with varying degrees of particulate backscattering and does not directly correspond to the bio-optical

conditions found in South African coastal waters; however, the application of the classification and algorithm blending framework to these data provide insight into the ability of the framework to resolve [Chl *a*] in bio-optical conditions that may occur outside of the region for which it was optimized.

Instead of supplying uncertainty estimates per water type for a single algorithm (as done in Moore et al., 2009), the current study offers a preliminary demonstration of blending the uncertainties from a selection of different algorithms per OWT. The MARE from the combined southern Benguela and MERMAID datasets were used to create the blended uncertainty product (results are shown in the Appendix in table B.1); each OWT is assigned an uncertainty value based on the performance of the assigned Chl *a* algorithm (determined in section 3.2.6). The weighting factors are obtained by normalizing the class memberships by the total membership sum, according to the following equation:

$$w_c = \frac{f_c}{\sum_1^x f_c} \tag{2.22}$$

where w_c is the normalized weighting factor for water type c, f_c is the fuzzy membership to water type c. Note that unlike the algorithm blending procedure in equation 2.20, the error weighting procedure has a different weight per OWT and not just per algorithm. The blended uncertainty product is then determined as follows:

$$Err_{blend} = \sum_1^x w_c Err_c \tag{2.23}$$

where Err_c is the uncertainty associated with the OWT from its assigned Chl *a* algorithm, w_c is the normalized weighting factor for water type c, and x is the number of OWTs. During the blending procedure it was assumed that the satellite reflectance vector had membership to at least one of the defined water types. If this is not the case the pixel is given a null value.

2.4 Synopsis

The objective of this chapter was to introduce the concept of optical water type classification of satellite data as a means of applying water type appropriate Chl *a* algorithms and blending the resulting products; this method promotes the likelihood that the causal IOP ranges of each reflectance spectrum falls within the operational limits of the algorithm that it is assigned to. The data, methods and techniques provided in this chapter represent a detailed description of the various data sources (*in situ*, synthetic, and satellite) that were systematically added in the stepwise development of a regionally representative clustering database and resulting optical water types. Lastly a variety of Chl *a* algorithms were described, as well as the statistical methods used to assign algorithms for each optical water type class.

The following chapter will include detailed results and discussion of the clustering methodology, optical water types, and the statistical performance and recommendations of the Chl *a* algorithms for each optical water type.

Chapter 3

Clustering results and Chlorophyll *a* algorithm selection

3.1 Introduction and chapter layout

Within the framework of the current study, the aim of the data clustering procedure is to produce physically or ecologically meaningful groups of data (clusters) that are representative of the coastal waters of South Africa. Although validity functions are available to guide the clustering process by providing quantitative evaluations of the structures imposed on the data, these measures do not guarantee ideal data groupings for the desired applications. The primary operational implementation of these clusters is satellite (MERIS) image classification as a means for Chl *a* algorithm application and blending. Due to the strong focus on the ability to retrieve quantitative [Chl *a*], a determining factor in assessing the appropriateness of cluster groupings is the ability of a cluster set to resolve distinct ranges of phytoplankton biomass, whilst still retaining the highest possible representativeness across a satellite image. Secondary applications for the clusters include informative qualitative characterization of the bio-optical provinces within a satellite image directly from R_{rs} without the need for a derived [Chl *a*] product, as well as the direct identification of optically distinct water types as stand-alone products, e.g. oligotrophic waters, harmful algal blooms, or Case 2 waters with specific algorithm requirements.

This chapter will describe the clustering results achieved throughout the systematic database expansion and modification process (as outlined in figure 2.1), starting with the original *in situ* dataset and followed by the consecutive augmentation with the synthetic and satellite data; EOF analysis of the datasets provides additional insight into the bio-optical drivers inherent to the various data sources. The classes are deconstructed

according to data source contributions in addition to class-specific EOF analysis to aid with appropriate Chl *a* algorithm selection. Lastly a class-specific error and uncertainty analysis of a range of Chl *a* algorithms indicates the most appropriate algorithms for the blending procedure, which is subsequently tested on coincidently measured satellite and *in situ* data matchups.

3.2 Results

3.2.1 Dataset construction and cluster selection process

The clusters resulting from each of the eight clustering steps in the database expansion and modification process can be seen in figures 3.1 - 3.8; these figures are grouped together below on pages 43 - 50. The clusters represented in each figure are not specifically ordered (e.g. according to [Chl *a*]). The input databases used for each of the clustering steps are described in figure 2.2. Each figure shows the number of clusters, the mean cluster spectra, the mean [Chl *a*], in addition to the four validity functions[1] (outlined in equations 2.4 to 2.7 in section 2.3.2) that were used as a guide to choose the appropriate number of clusters (c) for each step. However the choice for c often erred toward smaller values than indicated; an example of this can be seen in step[2] 1 (figure 3.1) where the validity functions indicated $c = 11$, whilst the second best choice ($c = 4$) was preferable as it provided larger clusters which tend to result in better satellite imagery classification success.

The mean classification success $\bar{A}$ (as defined per image by the A metric in equation 2.10) of each step per region, can be seen in table 3.1. In each step the clustering success was greatest in the west coast, while the east coast always showed the poorest performance; regional clustering success is discussed in more detail in chapter 4. There was an increase in cluster success from datasets 1 to 3 (S1 to S3) for all regions; a possible reason for this may be the increase in cluster sizes which in turn causes the covariance matrices of each cluster to be more inclusive.

The general cluster structure and mean shapes resulting from S2 (figure 3.2) remained very similar to those from S1, with only the addition of one extra cluster (C1) consisting of spectra representing very low biomass waters with a mean [Chl *a*] of 0.09 mg m^{-3}. After the addition of the synthetic Case 2 spectra in S3 (figure 3.3) the cluster selection was mostly skewed towards the data with high amounts of backscatter and reflectance peaks of >0.01 sr^{-1} between 480 and 580 nm. Approximately 72 % of the data were present in the first three clusters, while the other three each contained less spectra and elevated

[1]The desired results for validity functions are *PC* close to one, and *CE*, *SC*, and *S* close to zero.

[2]This section uses abbreviations to represent specific clusters and clustering steps during the database expansion and modification procedure. For instance, S2C3 would represent cluster three (or C3) from the second clustering step (or S2), whilst the addition of either M or E indicates the *moderate* or *elevated* groupings respectively, where applicable; S5C3M would thus represent cluster three from the moderate reflectance group in clustering step five.

Table 3.1: Table showing the details of the eight clustering steps, with the number of chosen clusters and available spectra. Also included are the mean ($\bar{A}$) ± standard deviation of the average total class membership sum of the ten test images for the west, south, and east coast domains respectively. See figure 2.2 for a description of the datasets.

Clustering step	**Dataset**	**Clusters**	**Spectra**	**West coast $\bar{A}$**	**South coast $\bar{A}$**	**East coast $\bar{A}$**
1	1	4[a]	273	0.208 ± 0.158	0.080 ± 0.039	0.008 ± 0.008
2	2	5	623	0.270 ± 0.189	0.130 ± 0.065	0.024 ± 0.0156
3	3	6	1028	0.486 ± 0.142	0.357 ± 0.106	0.101 ± 0.078
4	4M + 4E	5M + 5E	625 + 403	0.249 ± 0.182	0.111 ± 0.060	0.021 ± 0.016
5	5M + 5E	6M + 5E	1125 + 803	0.783 ± 0.212	0.677 ± 0.158	0.458 ± 0.171
6	6	5	1928	0.838 ± 0.137	0.813 ± 0.111	0.487 ± 0.116
7	6	9	1928	0.757 ± 0.251	0.582 ± 0.154	0.448 ± 0.161
8	7	5	1928	0.0151 ± 0.0124	0.0021 ± 0.0017	0.001 ± 0.0006

[a] *Although there were 4 clusters, only 3 were used for classification as the 4th cluster contained too few spectra*

reflectance; the cluster with the overall lowest reflectance (S3C2) contained 42 % of the data. The variability in the magnitude of the spectra, particularly in C4 to C6, were much more pronounced than C1 to C3 (note the change in the y-scale between these two groups). C4, C5 and C6 also have relatively similar and low [Chl *a*], suggesting that their spectral shapes were primarily affected by optically significant constituents other than phytoplankton. S3C2 contained a very high number of spectra and a [Cha *a*] range of 0.03 to 309 mg m^{-3}, which could ideally be grouped into more ecologically meaningful sub-groups. Previous studies have shown that the Euclidean distance performs well when the dataset has compact or isolated clusters (Mao and Jain, 1996), but that there is a tendency for the largest-scaled features to dominate the others (Jain et al., 1999); therefore the spectra from S3C2 were grouped together since they have lower reflectance which are relatively more similar to each other compared to the spectra within clusters 4 to 6.

After S3 the clustering database was divided into a moderate and elevated R_{rs} database. The terms *moderate* and *elevated* are used as a general description and are not a finite optical characterization of each group. The criteria placed spectra that showed either $R_{rs}(560) > 0.009$ sr^{-1}, or $R_{rs}(490) > 0.009$ sr^{-1} together with $R_{rs}(490)/R_{rs}(443) > 1$, in the elevated reflectance database, whilst the rest were added to the moderate reflectance database. The 560 nm region often displays a peak in reflectance, either from a minima in Chl *a* absorption combined with decreased backscattering from pure water or increased $b_{b\phi}$ in phytoplankton dominated waters, or from a notable inorganic particulate backscattering and/or CDOM absorption signal, or a combination of the two. The other criteria were based around the spectral shapes of S1C2 and S2C5. A R_{rs} value of 0.009 sr^{-1} was chosen as the cut-off for both 490 and 560 nm, as the data that had reflectance below this value in the green spectral region generally tended to display meaningful variability in spectral shape with changes in [Chl *a*]. The division allowed the separate clustering of spectra with generally low reflectance in the green to insure that subtle variations in spectral shape, particularly in phytoplankton dominated waters with [Chl *a*] between approximately 0.5 and 50 mg m^{-3}, would be accounted for. Clustering success decreased in S4 (figure 3.4) after the database separation. A possible reason for this may be that the clusters had fewer members, while the spectral ranges of the classes were narrower leading to less inclusive covariance matrices. However, the moderate reflectance classes were more ecologically meaningful and showed clear changes in mean spectral shape with increasing [Chl *a*].

After the addition of satellite data the mean classification success in S5 improved (based on the average success metric $\bar{A}$ in table 3.1) by a factor of 3.1 - 22.9 between S4 and S5, and 1.6 - 4.5 between S3 and S5 respectively; the most improvement was seen for the east coast. Possible reasons for the higher classification success over all regions may be the general increase in cluster size as a direct result of additional input data. Another reason may be the addition of more representative data, specifically the low biomass and elevated reflectance data represented by S5C2M and S5C7E respectively; these clusters could be particularly representative of the near-oligotrophic offshore waters and the turbid nearshore conditions often found along the east coast (this will be discussed in more detail in the next chapter). Whilst the elevated reflectance clusters in figure 3.5 maintained similar shapes to S4, the moderate reflectance group was rearranged and gained an additional low to medium biomass cluster. As in the previous step the moderate reflectance spectra were also naturally grouped according to [Chl *a*].

To further demonstrate the effects that the clustering database can have on the focus of the FCM clustering algorithm, S6 (figure 3.6) and S7 (figure 3.7) was performed on dataset 6 (all available data without a moderate and elevated reflectance database separation). The highest clustering success for the west and south coast was achieved in S6, possibly due to the large amount of data in each cluster. However, similar to S3 (figure 3.3), the clusters were skewed toward the higher reflectance data and most of the medium to high biomass data (814 spectra) were grouped together (S6C2) with [Chl *a*] range of between 0.1 and 464 mg m^{-3}, which is not meaningful with regard to phytoplankton focus. When increasing the number of clusters used by the algorithm, as done in S7 (figure 3.7), the clustering algorithm firstly focusses on the elevated reflectances and actually produces similar clusters (S7C5-9) to step 5 (S5C8E-11E). However the resulting clusters result in lower $\bar{A}$ than step 5; a possible reason for this is that they do not appear to adequately represent the moderate biomass range for phytoplankton-dominated waters ($\approx$ 2 - 20 mg m^{-3}).

The dataset of integral normalized reflectance, S8 (figure 3.8), showed very poor clustering success. A possible reason for this may include the narrow covariance matrices formed by the clusters. Another potential problem could stem from the amounts of variables (9 wavelengths) used during classification, which could lead to added points of distance; both Vantrepotte et al. (2012) and Mélin and Vantrepotte (2015) applied this method to SeaWiFS data and hence only had 6 weighting points, which may have lead to better classification success.

Ultimately S5 (utilizing dataset 5) presented the most distinct and ecologically meaningful cluster structures in terms of phytoplankton biomass groupings, with adequate clustering success over the test datasets. The average spectral shapes of the classes from dataset 5 resembled the shapes of the clusters from the original *in situ* dataset (S1C1 $\approx$ S5C3M; S1C3 $\approx$ S5C5M; S1C4 $\approx$ S5C4M; see figures 3.1 and 3.5); thus the clusters resulting from dataset 5 would provide improved classification of optical water types found in the original *in situ* data, with the addition of a selection of missing or distinct water types. The clusters from S5 will be used for further analysis; the individual clusters from this step will furthermore be referred to as classes or optical water types (OWTs).

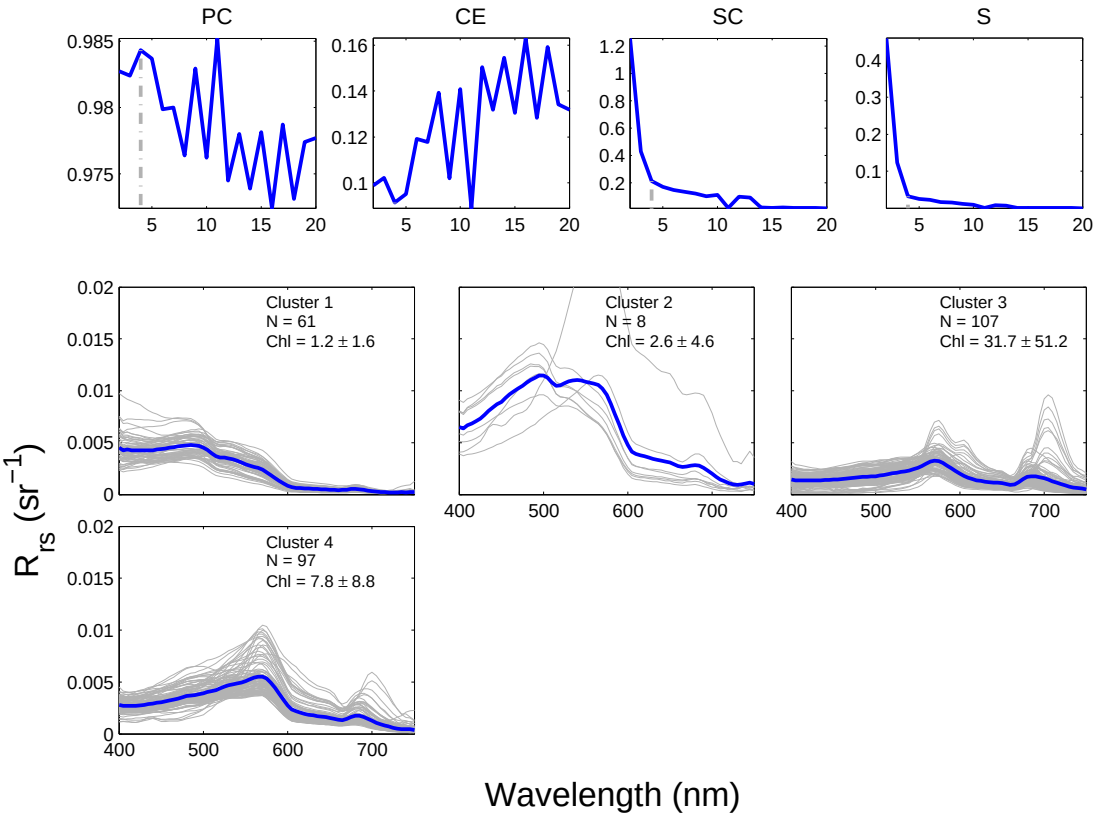


Figure 3.1: Clustering results from step 1, using dataset 1 (consisting of solely *in situ* data) with four clusters. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of four clusters (shown by the dashed lines) for the dataset. Also shown are the four resulting clusters. The legend of each plot shows the number of spectra N and the mean $\pm$ standard deviation of the [Chl a] per cluster. The blue lines in the R_{rs} panels represent the cluster means. (*Note: Cluster 2 was not used for image classification as it contained too few data*)

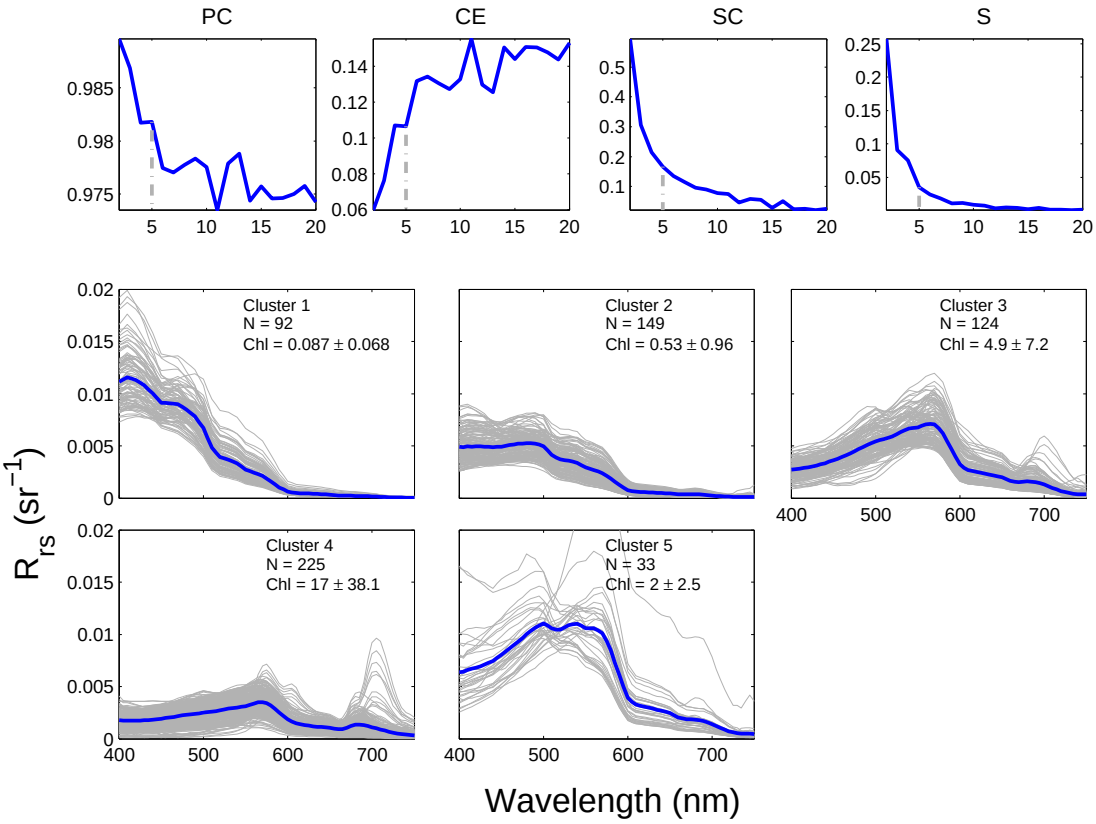


Figure 3.2: Clustering results from step 2, using dataset 2 (consisting of *in situ* and IOCCG synthetic data) with five clusters. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of five clusters (shown by the dashed lines) for the dataset. Also shown are the five resulting clusters. The legend of each plot shows the number of spectra N and the mean $\pm$ standard deviation of the [Chl a] per cluster. The blue lines in the R_{rs} panels represent the cluster means.

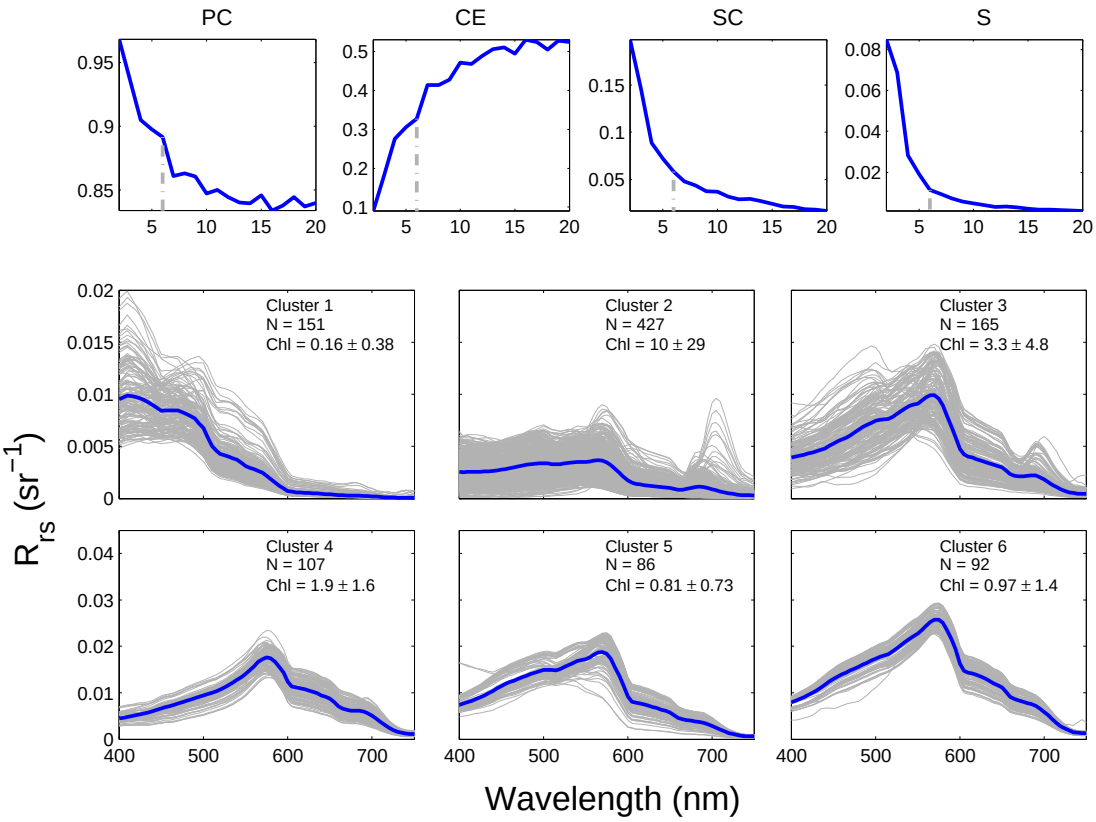


Figure 3.3: Clustering results from step 3, using dataset 3 (consisting of *in situ*, IOCCG and Case 2 synthetic data) with six clusters. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of six clusters (shown by the dashed lines) for the dataset. Also shown are the six resulting clusters. The legend of each plot shows the number of spectra N and the mean $\pm$ standard deviation of the [Chl a] per cluster. The blue lines in the R_{rs} panels represent the cluster means.

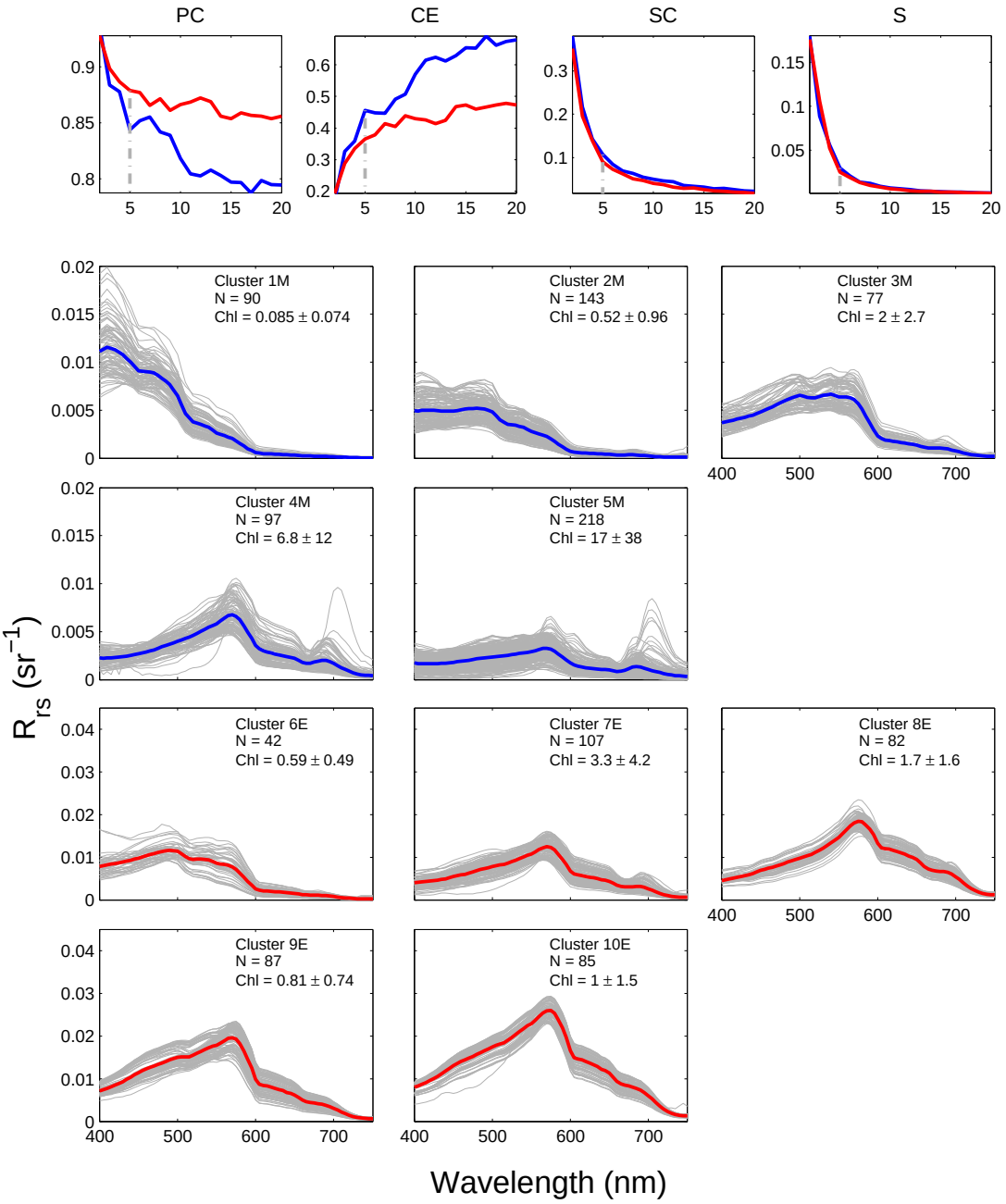


Figure 3.4: Clustering results from step 4, using dataset 4 (consisting of *in situ*, IOCCG and Case 2 synthetic data, modified by a database separation) with five clusters each for both the moderate and elevated reflectance groups. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of five clusters (shown by the dashed lines) for each dataset. The blue represents the moderate reflectance clusters, whilst red represents the elevated reflectance clusters. Also shown are the ten resulting clusters. The legend of each plot shows the number of spectra N and the mean $\pm$ standard deviation of the [Chl a] per cluster. The coloured lines in the R_{rs} panels represent the cluster means.

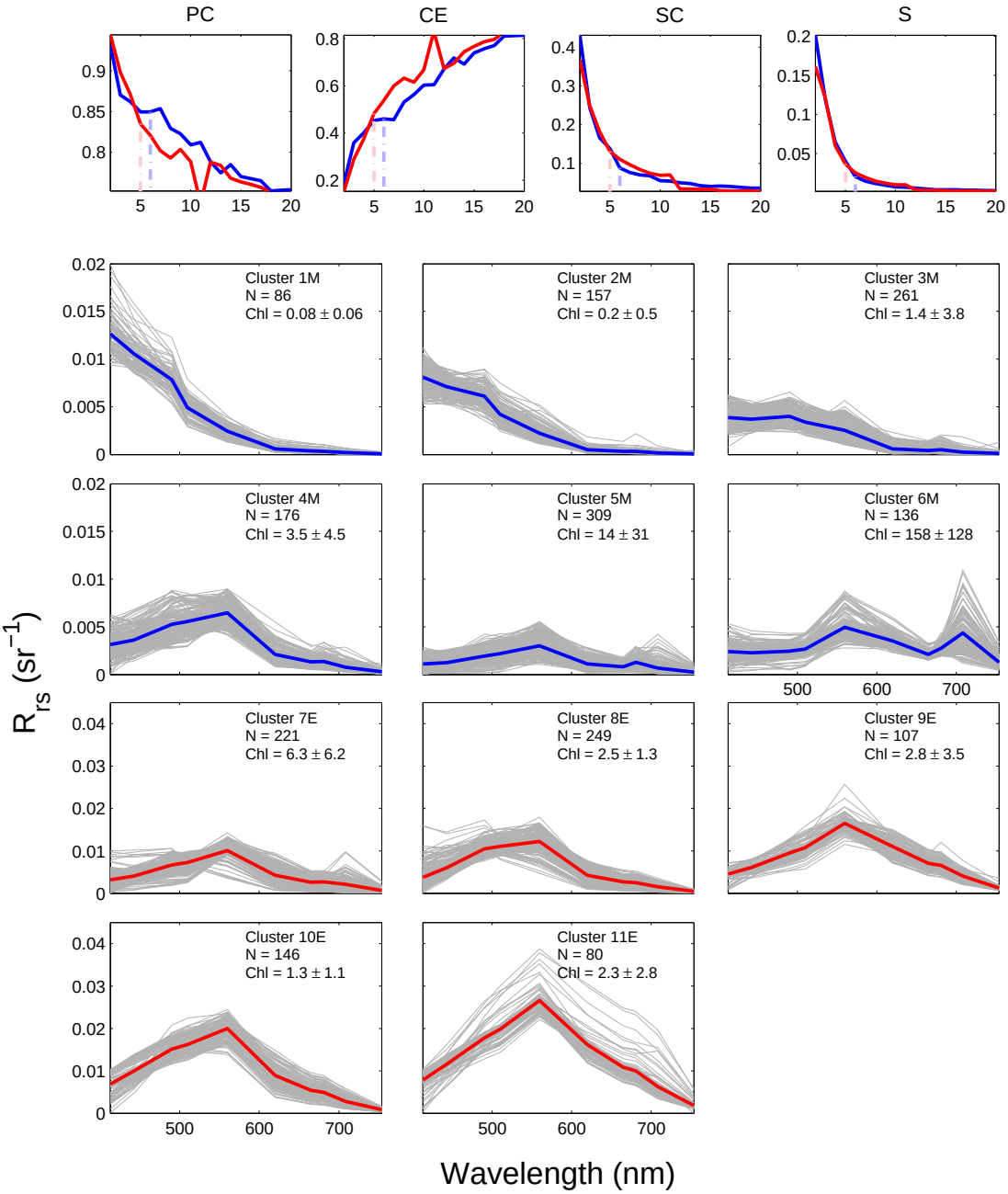


Figure 3.5: Clustering results from step 5, using dataset 5 (consisting of resampled *in situ*, IOCCG and Case 2 synthetic data, and extracted MERIS data, modified by a database separation) with six and five clusters for the moderate and elevated reflectance groups respectively. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of six and five clusters (shown by the dashed lines) for each dataset respectively. The blue represents the moderate reflectance clusters, whilst red represents the elevated reflectance clusters. Also shown are the eleven resulting clusters. The legend of each plot shows the number of spectra N and the mean $\pm$ standard deviation of the [Chl a] per cluster. The coloured lines in the R_{rs} panels represent the cluster means.

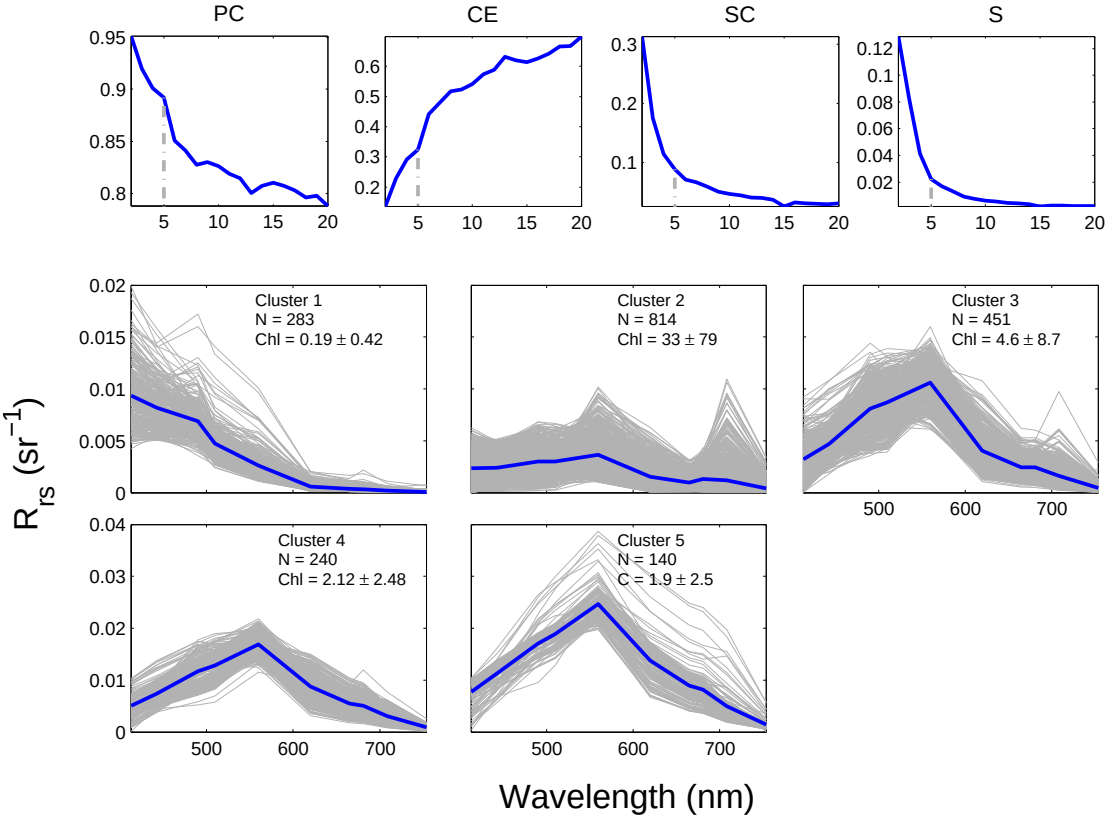


Figure 3.6: Clustering results from step 6, using dataset 6 (consisting of resampled *in situ*, IOCCG and Case 2 synthetic data, and extracted MERIS data without a database separation) with five clusters. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of five clusters (shown by the dashed lines) for each dataset. Also shown are the five resulting clusters. The legend of each plot shows the number of spectra N and the mean $\pm$ standard deviation of the [Chl a] per cluster. The blue lines in the R_{rs} panels represent the cluster means.

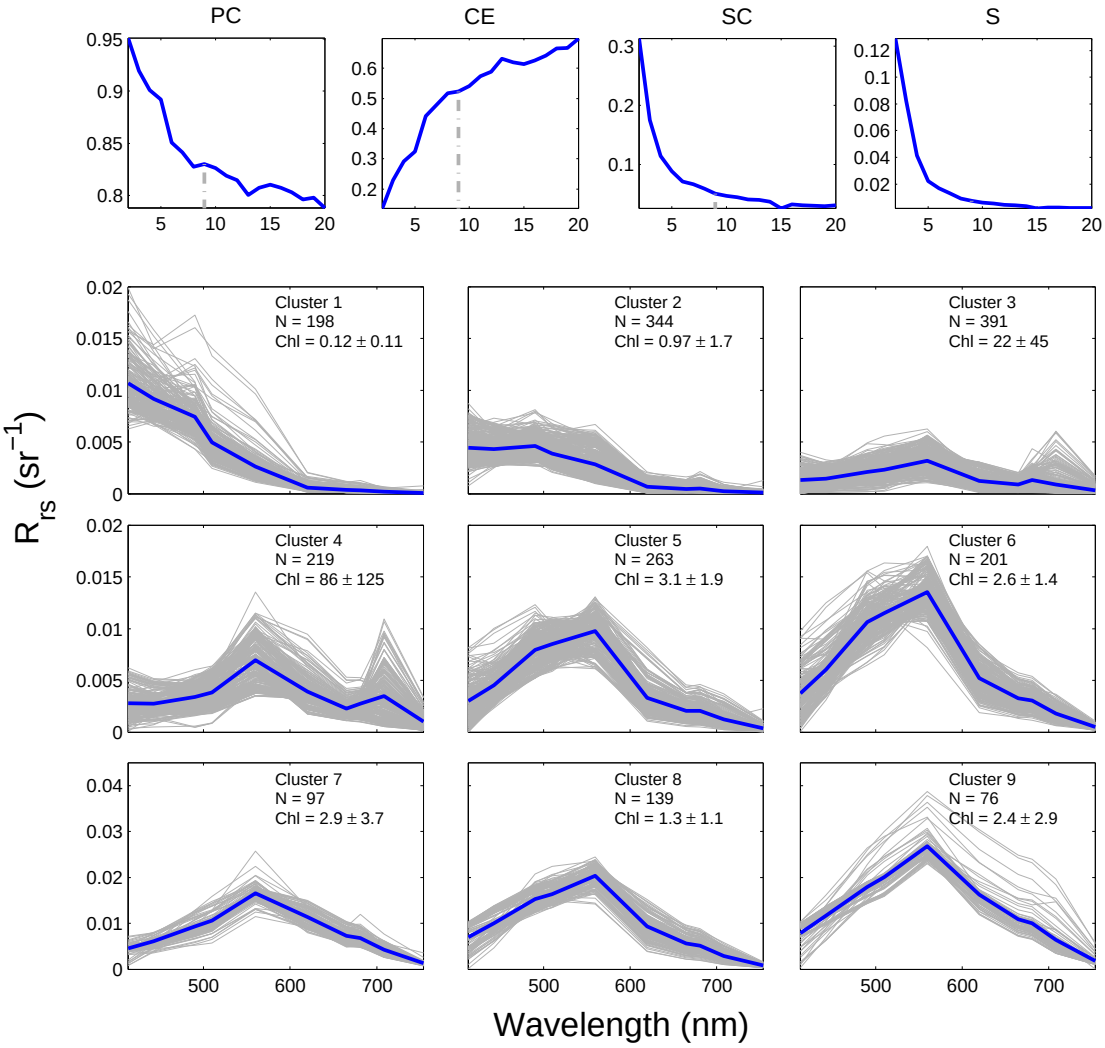


Figure 3.7: Clustering results from step 7, using dataset 6 (consisting of resampled *in situ*, IOCCG and Case 2 synthetic data, and extracted MERIS data without a database separation) with nine clusters. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of nine clusters (shown by the dashed lines) for each dataset. Also shown are the nine resulting clusters. The legend of each plot shows the number of spectra N and the mean $\pm$ standard deviation of the [Chl a] per cluster. The coloured lines in the R_{rs} panels represent the cluster means.

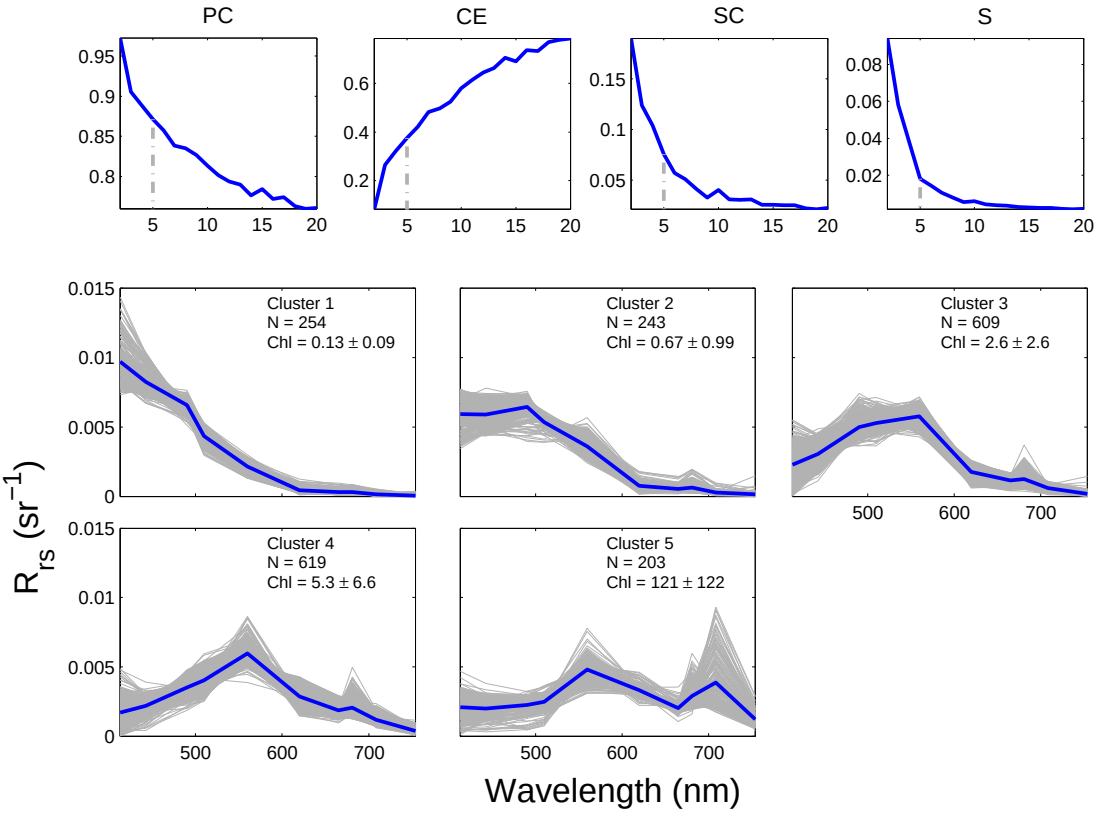


Figure 3.8: Clustering results from step 8, using dataset 7 (consisting of normalized *in situ*, IOCCG and Case 2 synthetic data, and extracted MERIS data) with five clusters. The top row of panels show the validity functions (Partition Coefficient PC, Classification Entropy CE, Partition Index SC and Separation Index S) which indicated the use of five clusters (shown by the dashed lines) for each dataset. Also shown are the five resulting clusters. The legend of each plot shows the number of spectra N and the mean ± standard deviation of the [Chl a] per cluster. The blue lines in the R_{rs} panels represent the cluster means.

3.2.2 Analysis of the optical variability within the clustering datasets

EOF analysis (as outlined in section 2.3.3) was applied to each of the clustering datasets (as described in figure 2.2) as an exploratory assessment of the clustering methods; the results provided insight into the bio-optical variability introduced by the addition of synthetic and satellite data, as well as the modifications such as database separation and normalization. The analyses were also performed on the final OWT classes to uncover the underlying bio-optical causality of each class. The EOF analysis of the first four clustering datasets can be seen in figure 3.9.

The first EOF mode ($M1$) of dataset 1 accounts for 64 % of the total R_{rs} variance. $M1$ is positive between 400 and 650 nm with a significant positive correlation (r=0.48) to $b_{bs}(560)$, emphasizing the influence of non-algal particles on the variability in spectral amplitude at this wavelength range. $M1$ is also significantly inversely correlated to [Chl a] (r=-0.41), and to a slightly lesser extent to the phytoplankton IOPs; the optical effects of increased phytoplankton biomass roughly translates to a respective decrease and increase of the reflectance in the blue and near-IR spectral regions. $M2$ shows two peaks near 570 and 700 nm, as well as a depression at 665 nm at the Chl a absorption peak. The Chl a fluorescence emission peak is generally considered to be centred at 681 nm; however, at high phytoplankton biomass a peak shift occurs towards longer wavelengths due to a combination of increasing Chl a absorption, increased particulate backscatter and pure water absorption. The fact that $M2$ has significant positive correlations to the phytoplankton IOPs and [Chl a], suggests that this mode is predominantly governed by the optical properties of phytoplankton; Craig et al. (2012) attributed a mode with a similar shape to the "bulk oscillations in biomass concentration".

Dataset 2 was the result of the addition of 350 spectra with generally lower [Chl a] ranges (0.01 - 5 mg m^{-3}) than the *in situ* dataset (0.13 - 309 mg m^{-3}). The effects of this can be seen in the shape of $M1$, which now has a spectral shape similar to R_{rs} of low biomass Case 1 waters. The $M1$ spectrum was significantly inversely correlated to [Chl a] and all the IOPs, and was also negative between 550 and 750, suggesting that these constituents affected this spectral region oppositely to the blue wavelengths. 33 % of the variance in this dataset is explained by $M2$, which shows a strong positive correlation (r=0.79) to $b_{bs}(560)$ as well as a weak positive correlation to $a_{dg}(440)$ (r=0.42).

After the addition of 405 synthetic Case 2 spectra 86 % of the variance in dataset 3 can be explained by a mode that has a strong positive correlation (r=0.74) to $b_{bs}(560)$, and a weak positive correlation (r<0.38) to phytoplankton absorption. The shape of $M1$ generally resembles the mean shape of the dataset between 400 and 750 nm, indicating that the variance was mainly due to changes in spectral amplitude (Craig et al., 2012). The spectral shape of the $M1$ is also similar to that of Lubac and Loisel (2007) who collected their data in the English Channel and southern North Sea, areas that were affected by coastal processes such as river inputs and resuspension of sediments, as well as spring phytoplankton blooms; they considered the first mode to be dominantly controlled by particulate backscattering. $M2$ only represents 11 % of the total variance and shows the most significant (negative) correlation to the phytoplankton IOPs.

Once the datasets were divided into two groups based on the reflectance ratios in the blue and green, it was possible to examine the primary optical drivers of each group more clearly. The moderate reflectance group had similar spectral characteristics and modal spectral shapes to dataset 2, with $M1$ and $M2$ contributing approximately 8 % more and 10 % less than dataset 2 respectively; although $M2$ was still significantly correlated to both a_{dg} and b_{bs}, the correlation was also smaller than in dataset 2. Both datasets 2 and 4M show that the third mode is significantly positively correlated to [Chl a] and the phytoplankton IOPs. Although the clusters within the moderate reflectance database tended to group together well according to phytoplankton biomass concentrations, the ambiguous correlations between the dominant mode of variance and the water constituents and IOPs indicates that the optical variability within this database is not solely biologically driven.

The first mode in the elevated reflectance database was responsible for 79% of the variance; since the shape of $M1$ and the mean reflectance spectrum of the database is very similar and entirely positive, it is likely that the first mode is due to changes in spectral amplitude driven mainly by non-algal particulate backscatter and the phytoplankton IOPs. $M2$ shows inverse correlation to [Chl a] and the phytoplankton IOPs, and positive correlation to b_{bs}.

The EOF analysis of clustering datasets 5 to 7 can be seen in figure 3.10. Once the satellite R_{rs} data were added and the datasets were resampled to satellite wavelengths, the EOF loadings were larger than for the hyperspectral data. The modes for the moderate reflectance group maintained very similar shapes between datasets 4 and 5, although the

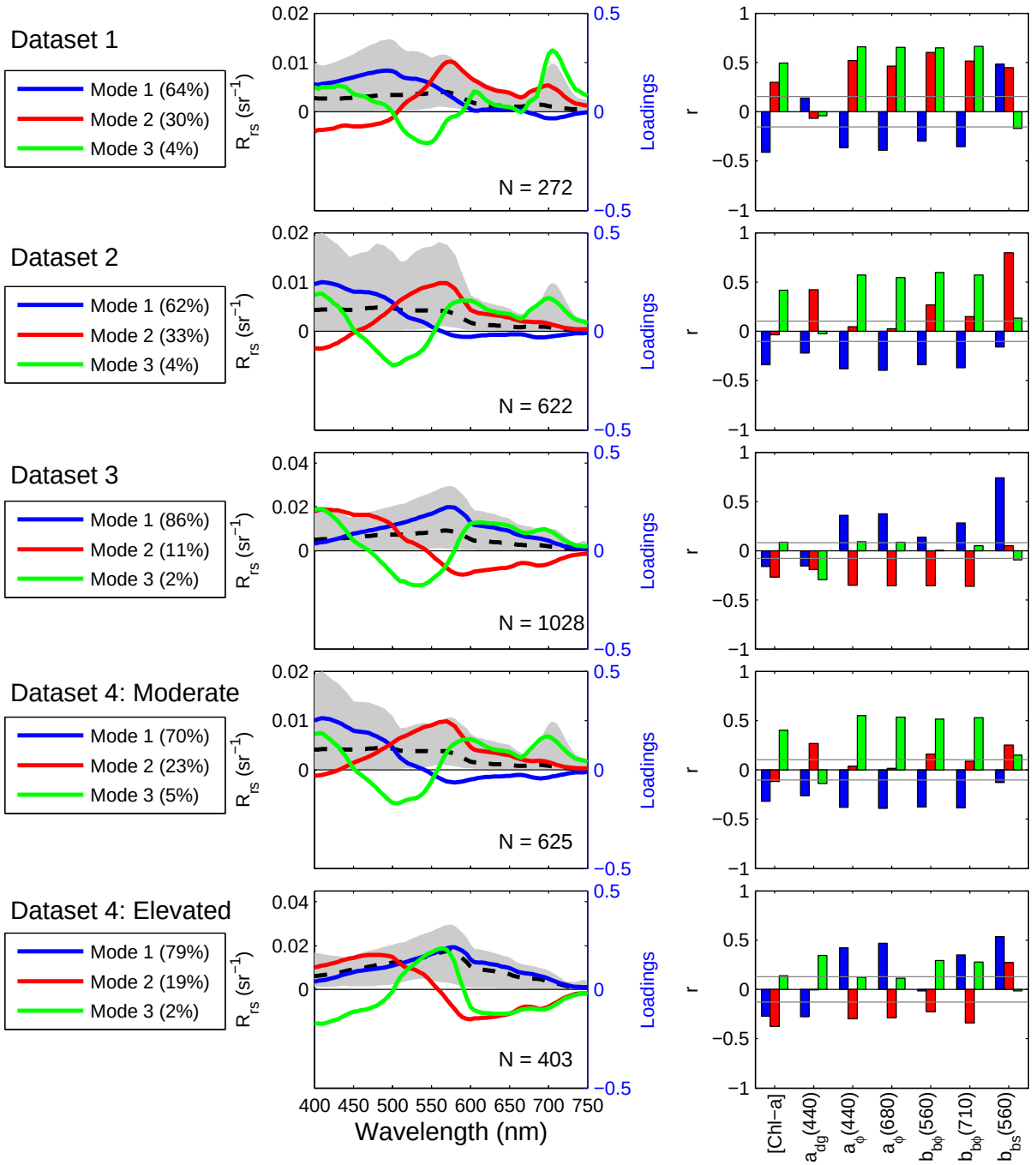


Figure 3.9: The top three EOF modes of clustering datasets 1 to 4 are displayed on the left, where the gray area represents the ranges of the R_{rs} spectra in each dataset. The correlation coefficients calculated between EOF amplitude factors and a selection of bio-optical parameters are displayed on the right, where the grey horizontal lines show the 99% confidence level for each plot.

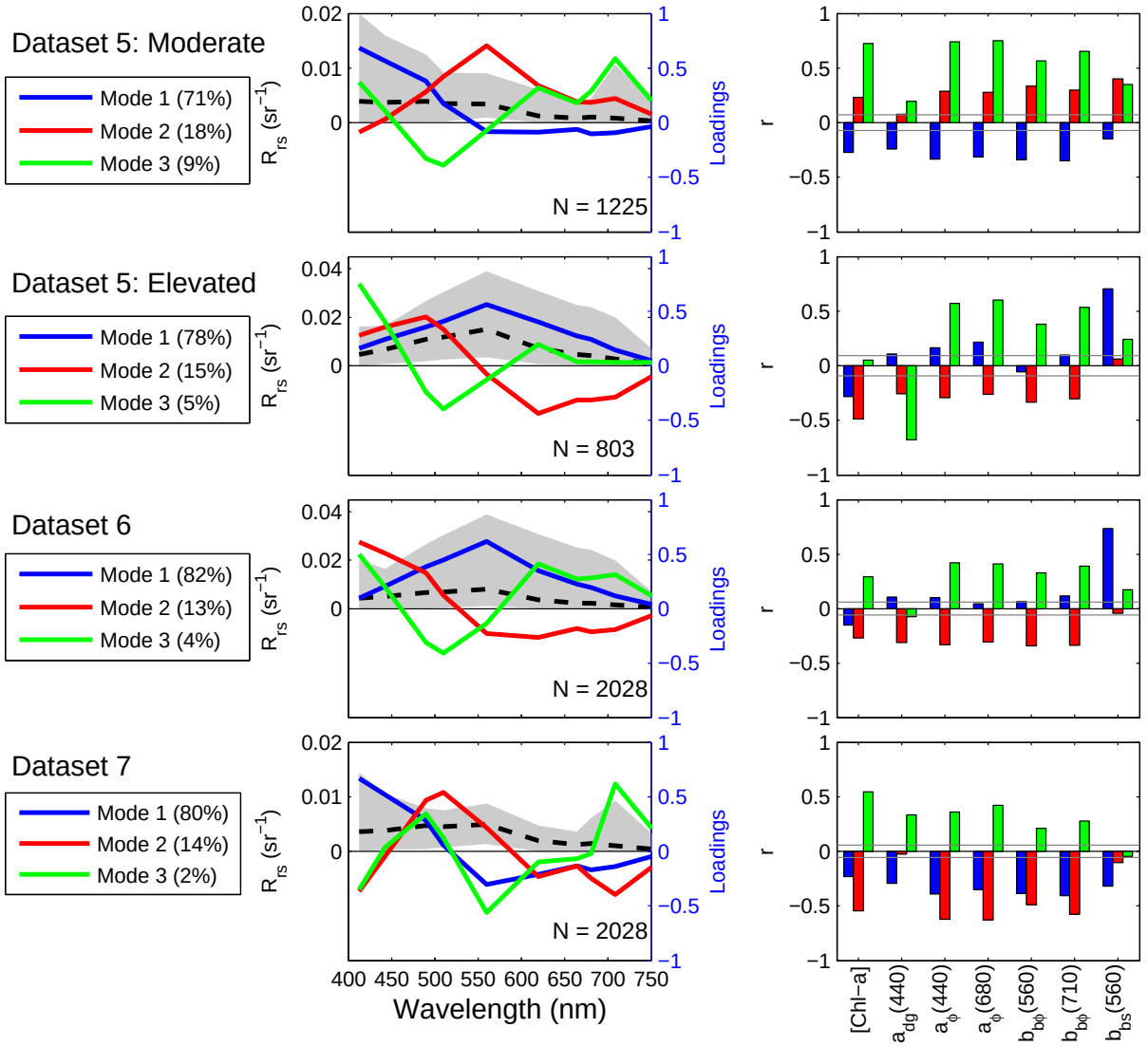


Figure 3.10: The top three EOF modes of clustering datasets 5 to 7 are displayed on the left, where the gray area represents the ranges of the R_{rs} spectra in each dataset. The correlation coefficients calculated between EOF amplitude factors and a selection of bio-optical parameters are displayed on the right, where the grey horizontal lines show the 99% confidence level for each plot.

second and third modes for dataset 5 were significantly and positively correlated to all the constituents. Similarly the shapes of $M1$ and $M2$ of the elevated reflectance database also maintained similar shapes between dataset 4 and 5; however $M1$ showed a stronger positive correlation to b_{bs}, whilst $M2$ showed a stronger negative correlation to [Chl a].

When using all available data, as done in dataset 6, 82 % of the variance of the dataset is mainly attributed to non-algal particulate backscattering (r=0.73). The relatively small second mode, similar to $M2$ from dataset 3, is negatively correlated to all the constituents other than b_{bs}. The normalization of the dataset, as done for dataset 7, produces a similar first mode of variance to dataset 5M, which is negatively affected by [Chl a] and all the IOP constituents. The second mode of variance in this dataset is also negatively correlated to [Chl a] and the phytoplankton IOPs.

Overall the EOF analyses demonstrate the efficacy of the database separation, as there is a clear differentiation between the bio-optical drivers between the dominantly phytoplankton controlled moderate reflectance database, and the elevated database where b_{bs} dominated the bio-optical causality. This separation facilitates an approach similar to integral normalization of a R_{rs} database, where the focal point of the clustering of the moderate reflectance group is driven by the spectral shape rather than the magnitude; the separation procedure can thus provide the benefit of the normalization approach (i.e. the focus on spectral shape) without sacrificing classification performance (i.e. being able to relate clusters directly to phytoplankton biomass).

3.2.3 The optical water type classes: data composition and assessment of class overlap

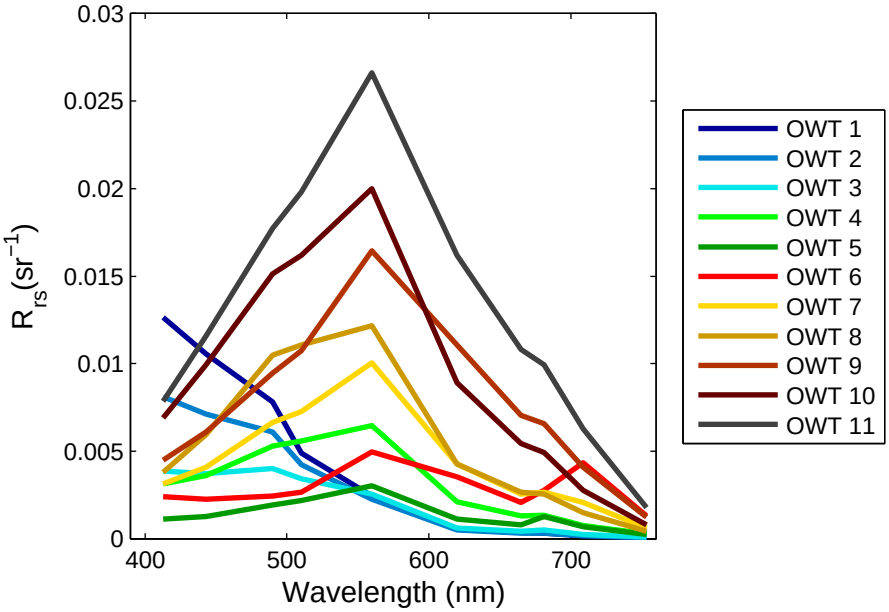
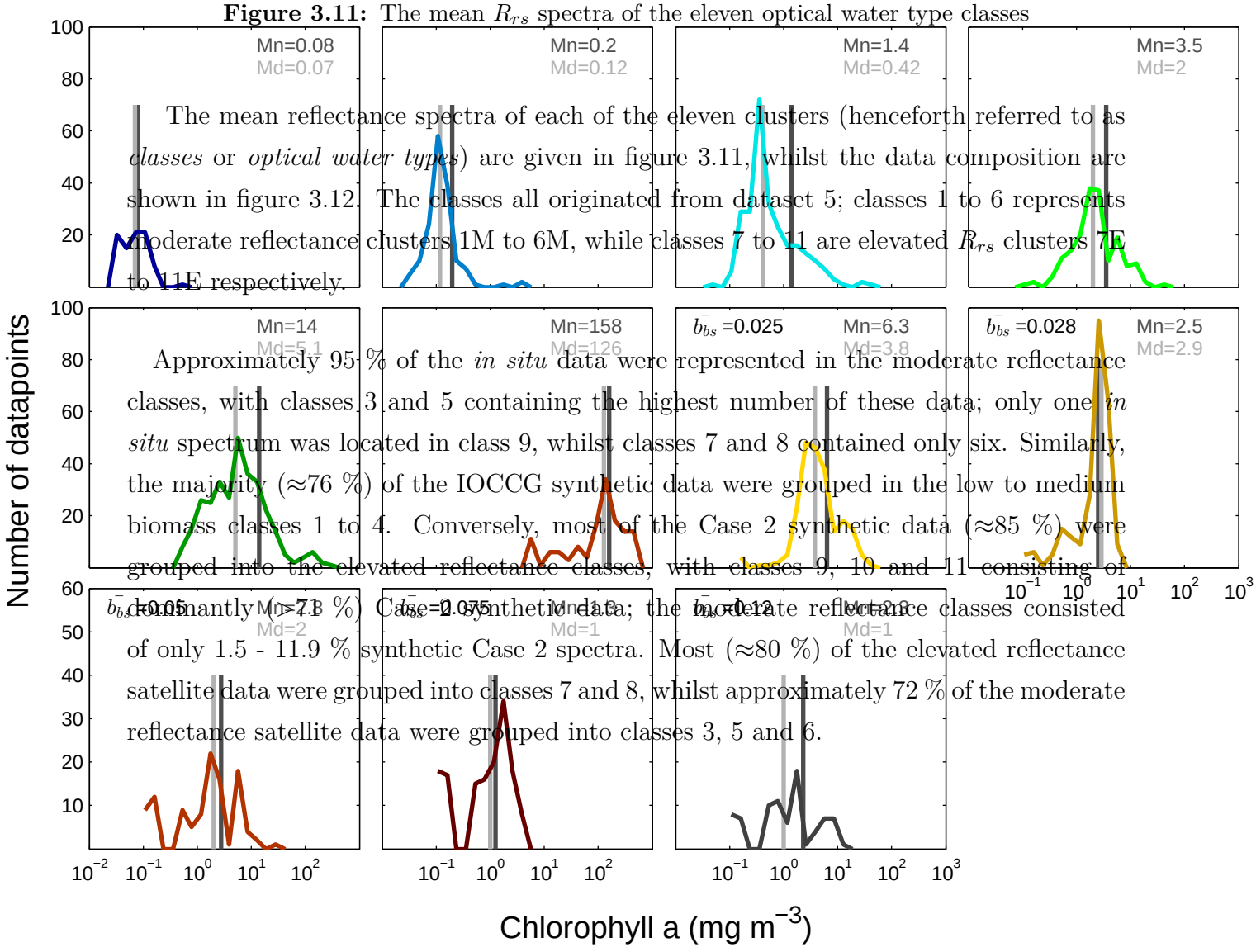


Figure 3.11: The mean R_{rs} spectra of the eleven optical water type classes

The mean reflectance spectra of each of the eleven clusters (henceforth referred to as *classes* or *optical water types*) are given in figure 3.11, whilst the data composition are shown in figure 3.12. The classes all originated from dataset 5; classes 1 to 6 represents moderate reflectance clusters 1M to 6M, while classes 7 to 11 are elevated R_{rs} clusters 7E to 11E respectively.

Approximately 95 % of the *in situ* data were represented in the moderate reflectance classes, with classes 3 and 5 containing the highest number of these data; only one *in situ* spectrum was located in class 9, whilst classes 7 and 8 contained only six. Similarly, the majority (≈76 %) of the IOCCG synthetic data were grouped in the low to medium biomass classes 1 to 4. Conversely, most of the Case 2 synthetic data (≈85 %) were grouped into the elevated reflectance classes, with classes 9, 10 and 11 consisting of predominantly (>71 %) Case 2 synthetic data; the moderate reflectance classes consisted of only 1.5 - 11.9 % synthetic Case 2 spectra. Most (≈80 %) of the elevated reflectance satellite data were grouped into classes 7 and 8, whilst approximately 72 % of the moderate reflectance satellite data were grouped into classes 3, 5 and 6.



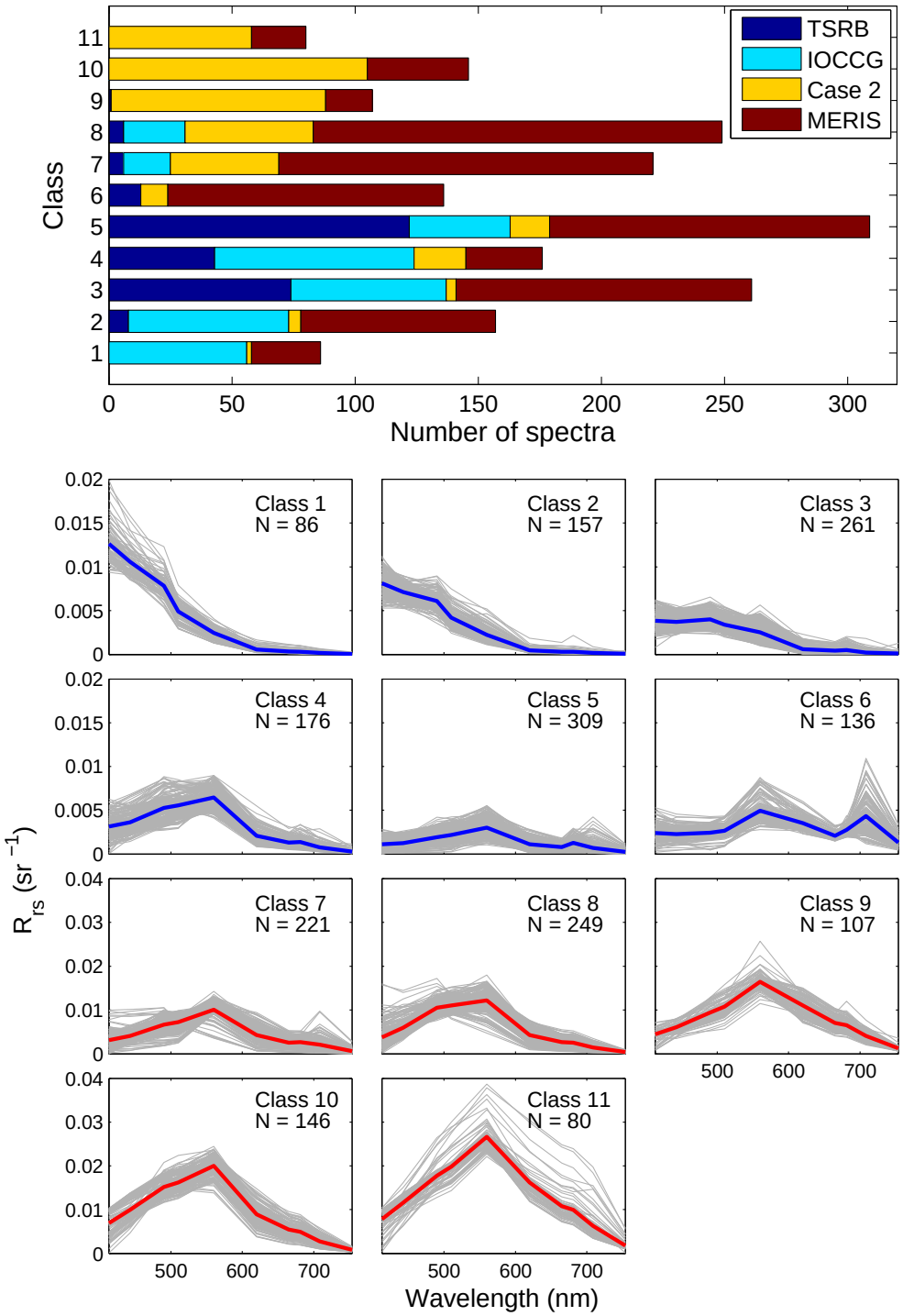


Figure 3.12: Top: a stacked bar graph showing the data type composition of the final classes. These included *in situ* data (TSRB), two types of synthetic data (IOCCG and Case 2) as well as extracted satellite data (MERIS). Bottom: the eleven final optical water type classes. The grey lines represent the individual spectra that make up each class, whilst the coloured lines show the class means. Classes 1 - 6 (blue) represent the moderate reflectance classes, whilst classes 7 - 11 represent the elevated reflectance classes.

Table 3.2: The mean membership of each of the eleven OWT classes to each of the eleven mean class spectra. Mean memberships <0.005 are not shown (replaced with -). The shaded region shows the OWTs from the moderate reflectance database, whilst the rest are from the elevated reflectance database. [Notation: M1 represents mean 1 whilst C1 represents the data grouped into class 1]

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
C1	**0.61**	0.12	-	-	-	-	-	-	-	-	-
C2	0.18	**0.64**	0.04	0.01	-	-	-	-	-	-	-
C3	-	0.08	**0.65**	0.19	0.04	-	-	-	-	-	-
C4	-	-	0.03	**0.63**	0.02	0.01	0.15	0.06	-	-	-
C5	-	-	0.04	0.10	**0.63**	0.05	-	-	-	-	-
C6	-	-	-	-	-	**0.63**	0.02	-	-	-	-
C7	-	-	-	0.04	-	-	**0.62**	0.24	0.03	-	-
C8	-	-	-	-	-	-	0.19	**0.63**	0.04	0.03	-
C9	-	-	-	-	-	-	-	0.01	**0.70**	0.03	-
C10	-	-	-	-	-	-	-	0.03	0.02	**0.65**	0.08
C11	-	-	-	-	-	-	-	-	-	0.08	**0.69**

The validity functions provide some indication of the amount of separation and overlap between the clusters within a dataset; however, since the resulting eleven (six and five) OWTs were developed independently of one another it is necessary to assess the extent of overlap that may exist between classes from the different datasets. Significant overlap between classes from different databases may require class re-configuration. In order to assess this potential overlap each of the clusters were classified individually with the eleven class means and covariance matrices; the resulting mean memberships are shown in table 3.2. Each of the OWTs have the highest membership to its own class mean, and all of these values are >0.60. It is evident that some overlap does exist between the OWTs from the two different databases, with class 4 showing a mean membership of approximately 0.15 to the mean of class 7; however the degree of overlap appears to be similar or smaller than the intra-database class overlap (e.g. between OWTs 7 and 8), suggesting that the class structures would not require further re-configuration.

3.2.4 The optical water type classes: analysis of causal bio-optical variability

EOF analysis was performed on each of the final classes to determine the causal optical properties contributing to the variance and shape of each class. The causal optical characteristics of the moderate reflectance classes can be seen in figure 3.13 and are described as follows:

- Class 1 represents typical near-oligotrophic waters with a very low [Chl a] (mean 0.08 mg m^{-3}, s.d. 0.06). Generally there would be very low concentrations of optically significant constituents in the water to affect the R_{rs} shape of this class and thus the dominant mode of variance has a shape similar to the scattering coefficient of pure water. $M1$ accounts for 69 % of the class variance and has only a significant negative correlation to $a_{dg}(440)$ (r=-0.41), which appears to control the magnitude of the R_{rs} in the blue. $M2$ has a positive shape between 440 and 750 nm and has a significant positive correlation to $a_{dg}(440)$, [Chl a], and a_{ϕ}.

- Class 2 shows a slight increase in [Chl a] compared to the first OWT (mean 0.2 mg m^{-3}, s.d. 0.5), although it still represents relatively low biomass conditions. Similar to class 1, $M1$ from class 2 maintains a spectral shape similar to pure water scattering, but represents a smaller portion of the total variance and has a significant negative correlation to both [Chl a] and $a_{dg}(440)$. The shape and amplitude of $M2$ is most likely controlled by $b_{b\phi}(560)$ and a_{dg}. $M3$ is negative in the green and red, and shows significant negative correlation to [Chl a]; it also shows a dip in the spectrum in the region of the Chl a fluorescence peak, which would indicate some influence of the Chl a fluorescence signal near 681 nm.

- Class 3 represents low to moderate [Chl a] (mean 1.4 mg m^{-3}, s.d. 3.8). The percentage contribution of the first two modes to the total variance is smaller compared to previous classes, which may be an indication of increased complexity of the variance causality of this OWT. $M1$ is positive in the blue and negative in the green and red, suggesting that the IOPs impact these spectral regions differently; $M1$ also shows significant negative correlations with [Chl a] and all the phytoplankton IOPs, which means that an increase in phytoplankton biomass would cause a decrease in reflectance in the blue, and an increase in reflectance in the green and red. $M2$ is positive with a peak in the green, and has significant positive correlation to $b_{b\phi}$ and b_{bs}, indicating these IOPs as controlling variables. Class 3 has a larger third

mode than the previous classes (11 %) and is most strongly and negatively correlated to $a_{dg}(440)$ (r=-0.52), but it is also affected by [Chl a] and the phytoplankton IOPs; the mode appears to be influenced by phytoplankton biomass, since it has a Chl a fluorescence peak at 681 nm and is negative between 470 and 550 nm where phytoplankton absorption increases with biomass.

- Class 4 represents moderate [Chl a] (mean 3.5 mg m^{-3}, s.d. 4.5). $M1$ is positive in the blue and negative between 560 and 750 nm and is negatively correlated to [Chl a], a_{dg} and $b_{b\phi}$. The shape and amplitude of $M2$ appears to be primarily influenced by a_{dg} and $b_{b\phi}$. Similar to class 3, the influence of increasing phytoplankton biomass is demonstrated by the significant positive correlation of $M3$ to [Chl a] and the phytoplankton IOPs, as well as the presence of a fluorescence peak at 681 nm.

- Class 5 contains the most *in situ* data and represents the high biomass waters (mean 14 mg m^{-3}, s.d. 31) prevalent in the southern Benguela. $M1$ is positive between 300 and 750 nm, and is weakly and negatively correlated to [Chl a] and a_{gd} which appear to have the dominant controlling influence in class 5. $M2$ shows a strong positive correlation to the phytoplankton IOPs and [Chl a]. The modal shape has a trough near 660 due to strong absorption by Chl a; it also has a Chl a fluorescence peak at 681 nm and a peak at 710 nm due to the combined effects of water absorption and phytoplankton backscatter in this region. $M3$ is significantly negatively correlated to a_{gd}, and positively to [Chl a] and a_{ϕ}; the shape of the mode is similar to that of classes 3 and 4, but with a larger negative region (480 to 620 nm) indicating stronger phytoplankton absorption.

- Class 6 consists primarily of extracted MERIS data and has the highest [Chl a] (mean 158 mg m^{-3}, s.d. 128). The first mode is negative between 400 and 600 nm, with a strong backscattering-induced peak at 710 nm. $M1$ is mainly influenced by a_{ϕ}, but also shows strong positive correlation to $b_{b\phi}$ and [Chl a], and slightly less to a_{gd} and b_{bs}. $M2$ has a positive spectral shape that is very similar to the class mean; the mode is only significantly correlated with $b_{bs}(560)$, indicating that the amplitude of this mode is affected by non-algal particulate backscattering. $M3$ is negative from 400 to 500 nm, and has a negative and positive correlation to [Chl a] and $b_{b\phi}$ respectively; this indicates that as biomass increases, blue wavelengths contain less signal while red and green wavelengths are increasingly dominated by phytoplankton scattering.

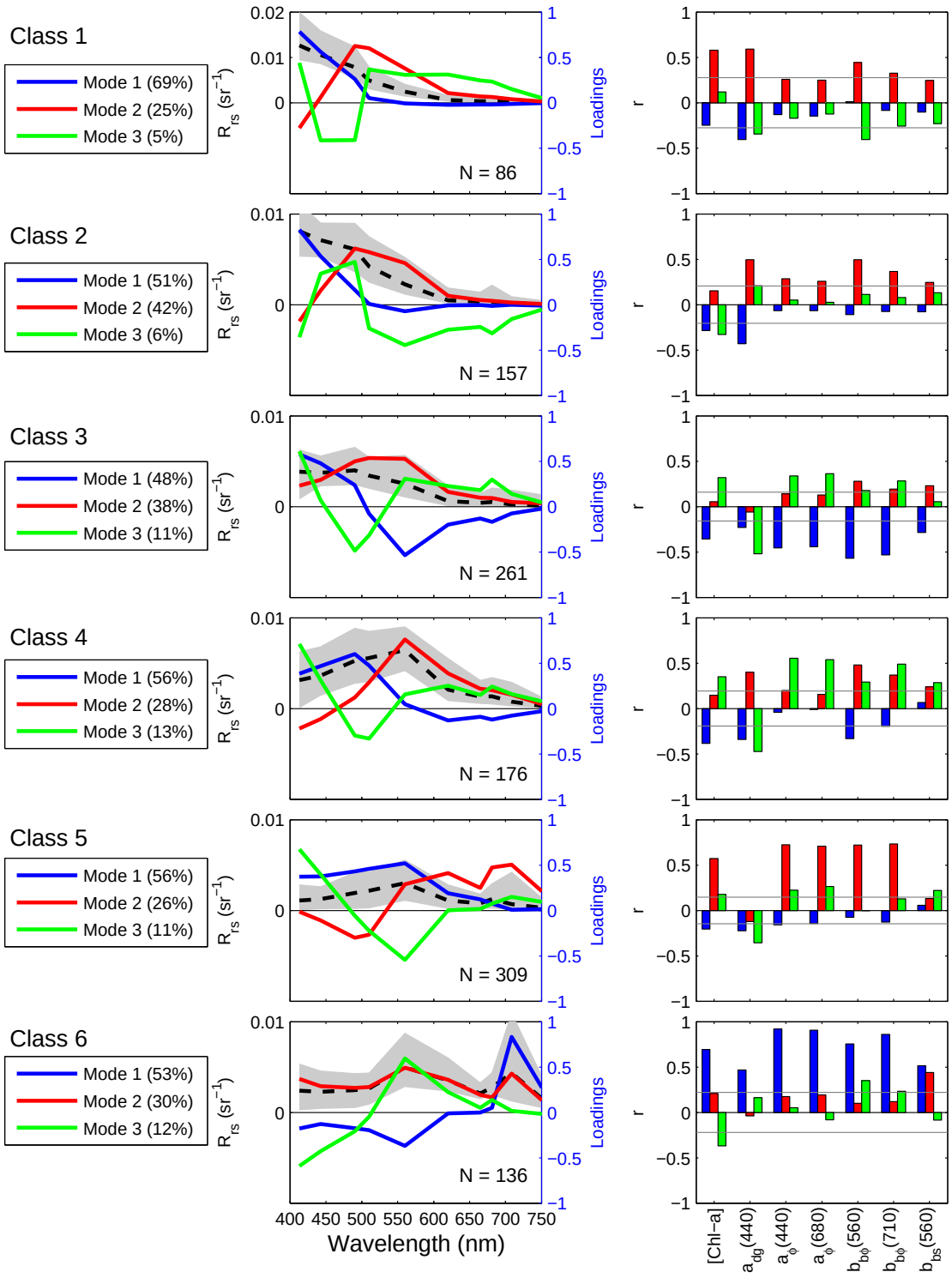


Figure 3.13: The top three EOF modes of the moderate reflectance classes 1 to 6 are displayed on the left, where the gray area represents the ranges of the R_{rs} spectra in each dataset. The correlation coefficients calculated between EOF amplitude factors and a selection of bio-optical parameters are displayed on the right, where the grey horizontal lines show the 99% confidence level for each plot.

The causal optical characteristics of the elevated reflectance classes can be seen in figure 3.14 and are described as follows:

- Class 7 is fairly constrained by the limitations imposed by the database separation, and resultantly it contains a variety of spectra that are grouped together by the FCM algorithm due to their relative similarity and lower reflectance compared with the rest of the elevated reflectance group. The class should be considered as a transition between the moderate and elevated reflectance datasets, since it may represent spectra of moderate [Chl a] (mean 6.3 mg m^{-3}, s.d. 6.2); this is demonstrated by the relatively strong negative correlation between $M1$ and [Chl a] and the phytoplankton IOPs. $M2$ is most strongly and negatively correlated with $a_{dg}(440)$, but also shows a weak positive correlation to [Chl a] and the phytoplankton IOPs; the mode also has a similar shape and causality to mode 3 from class 5. $M3$ is overall positive with the strongest correlation to b_{bs}.

- In class 8 the first mode of variance showed a strong negative correlation to $a_{dg}(440)$ and [Chl a]. $M2$ was completely positive, with weak positive correlations to the phytoplankton IOPs and $b_{bs}(560)$. The third mode was negatively correlated to phytoplankton IOPs and $b_{bs}(560)$, and positively to $a_{dg}(440)$.

- Both classes 9 and 10 contained primarily Case 2 synthetic spectra, whilst the dominant mode of variance $M1$ showed almost no significant correlation to any of the constituents; these classes may represent a mixture of water types where it may be difficult to separate the causal contributions of a_{dg}, b_{bs} and the phytoplankton IOPs to the spectral shape. Both classes had a relatively large second mode. $M2$ of class 9 was significantly, but negatively, correlated to the phytoplankton IOPs and positively correlated to $a_{dg}(440)$, whilst $M3$ showed primarily a strong negative correlation to $a_{dg}(440)$ (r=-0.8). $M2$ from class 10 was significantly correlated to $b_{bs}(560)$, but was also affected by all the other constituents, whilst $M3$ was strongly correlated to the absorbing IOPs. Class 10 has the lowest [Chl a] (mean 1.3 mg m^{-3}, s.d. 1.10) of the elevated reflectance classes.

- Class 11 is the smallest class and has the highest overall reflectance. $M1$ contains 74 % of the total variance and is significantly affected by all the constituents; the mode also shows the strongest correlation to $b_{bs}(560)$ out of all the elevated reflectance classes and represents the most turbid class. $M2$ only shows a significant (negative) correlation to [Chl a] and $a_{dg}(440)$. The small (5 %) $M3$ is influenced mainly by the phytoplankton IOPs. All the other classes from the elevated reflectance database

show relatively large contributions from modes 2 and 3 to the total variance, whilst these modes appear to show increasingly complex (or even unclear) causality.

Overall the aforementioned results compare favourably with those found in previous studies using these type of analyses (e.g Lubac and Loisel, 2007; Craig et al., 2012). Previous studies have performed EOF analysis on the entire reflectance datasets (Mueller, 1976; Toole and Siegel, 2001; Lubac and Loisel, 2007; Craig et al., 2012); however, assessing the dominant modes of variance within each class together with the potential causes of the variance provides a unique perspective on the bio-optical causality of each class, and could potentially aid the appropriate algorithm selection for the individual OWTs.

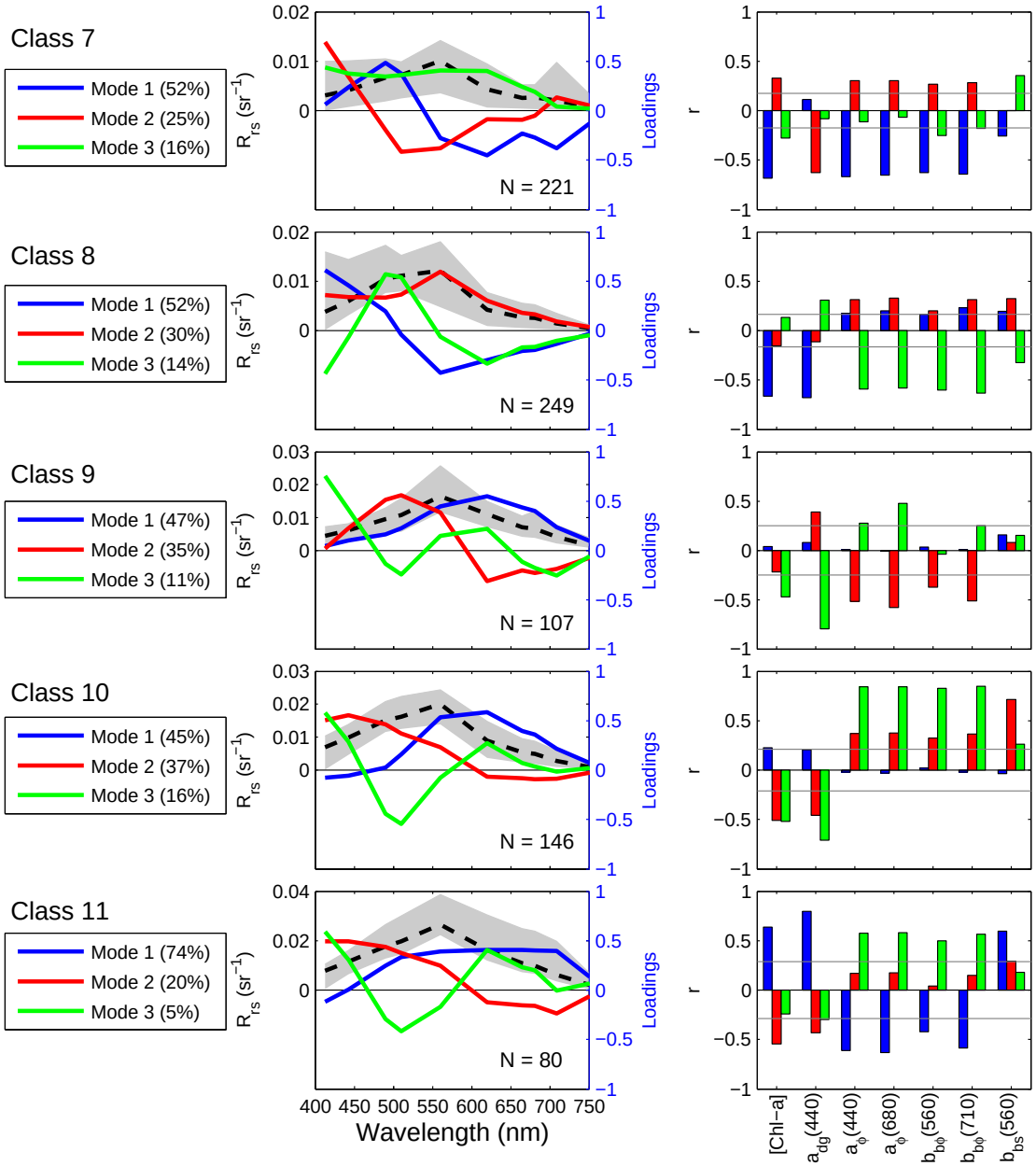


Figure 3.14: The top three EOF modes of the elevated reflectance classes 7 to 11 are displayed on the left, where the gray area represents the ranges of the R_{rs} spectra in each dataset. The correlation coefficients calculated between EOF amplitude factors and a selection of bio-optical parameters are displayed on the right, where the grey horizontal lines show the 99% confidence level for each plot.

3.2.5 Descriptive qualities of the optical water type classes

The moderate reflectance classes were arranged according to increasing [Chl *a*] as reflected by the clear increase in mean and median Chl *a* values from OWTs 1 to 6 in figure 3.15. The elevated reflectance OWTs 7 to 11 were generally arranged to represent increasing backscattering and turbidity and displayed overall lower mean and median [Chl *a*]. The [Chl *a*] distributions for classes 9 to 11 were more uneven and without a dominant peak, with frequency peaks that may be attributed to the synthetic data [Chl *a*] steps. The moderate reflectance classes (1 to 6), as well as class 7 and 8 of the elevated reflectance group, displayed a lognormal distribution for [Chl *a*] (figure 3.15) each spanning between 2 and 4 orders of magnitude; thus it is practical to log-transform the data prior to statistical analysis to allow balanced error assessments across these broad ranges.

The moderate reflectance classes display the optical characteristics associated with increasing phytoplankton biomass. The first two classes correspond to the classic Case 1 scenario where the shape of the water-leaving reflectance is primarily influenced by [Chl *a*] and co-varying degradation products such as a_{gd}; these classes likely represent near-oligotrophic waters dominated by small celled phytoplankton species. As the amount of phytoplankton in the water increases, the light that is absorbed by Chl *a* increases proportionally, leading to an incremental shift in the reflectance peak near 560 nm towards longer wavelengths (Dierssen et al., 2006; Robertson Lain et al., 2014); this effect was displayed in the change of spectral shape between classes 1 to 4. Increasing phytoplankton concentrations also typically lead to enhanced $b_{b\phi}$ manifesting as increased water-leaving radiance between 550 and 700 nm (Dierssen et al., 2006); these effects are generally tempered by increased absorption (Robertson Lain et al., 2014) as expressed in the overall low magnitude and mean reflectance of class 5.

Evers-King et al. (2014) found that the relative effect on R_{rs} from changes in cell size of phytoplankton becomes apparent in the green as biomass levels increase above 1-3 mg m^{-3} (i.e. in OWT4), when smaller cells sizes may lead to elevated R_{rs} in this spectral region; the previous section also showed that $b_{b\phi}$ plays an important role in determining the variance in the spectral R_{rs} of OWT4. Keeping this in mind it is plausible that OWT4 corresponds to the early stages of succession (thought to be small diatoms according to Margalef, 1962 cited in Wasmund et al., 2014) found within maturing upwelled water that usually consists of biomass ranges of 1-20 mg m^{-3} (Barlow, 1982; Barlow et al., 2005). Within this context the overall low magnitude and mean reflectance of OWT5 may be an indication of larger cells at higher biomass levels; within the upwelling scenario this

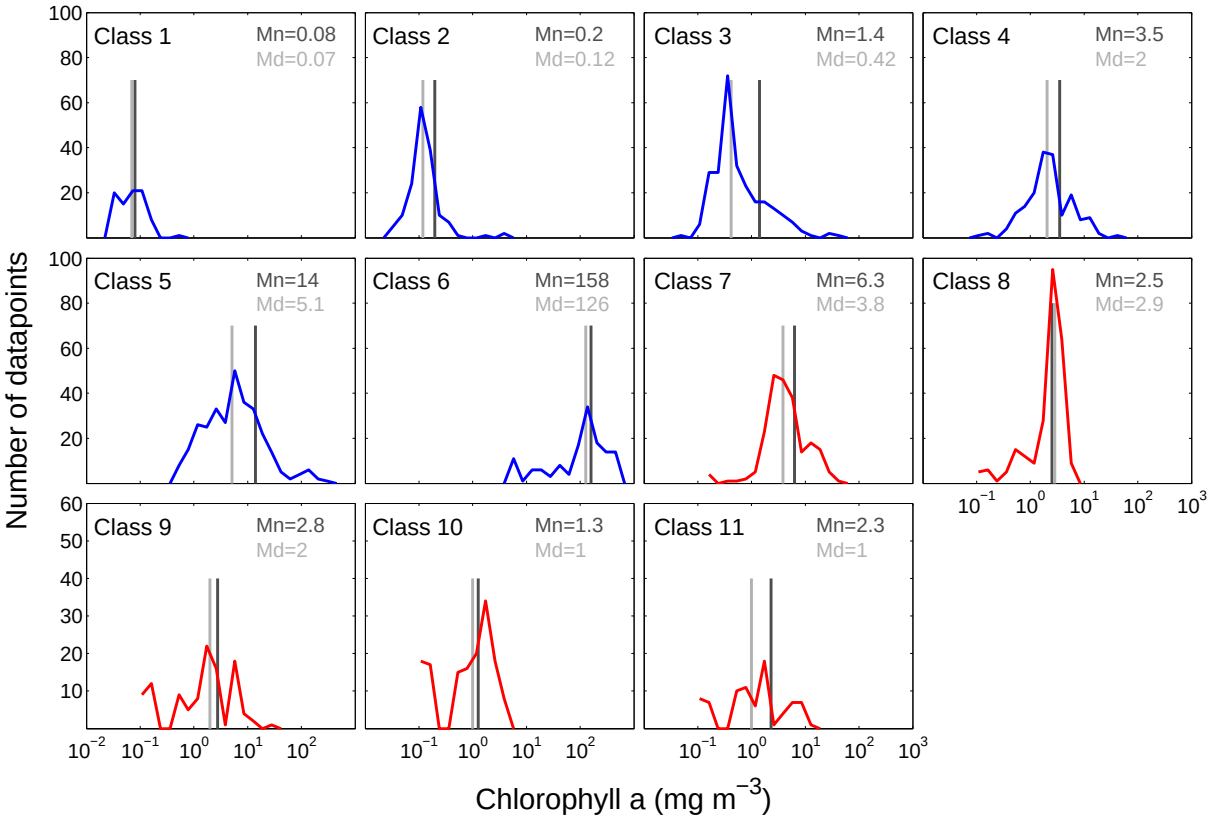


Figure 3.15: Histograms showing the [Chl *a*] ranges and frequency of occurrence for each of the eleven classes. Each class mean (dark grey line) and median (light grey line) are shown for reference. These plots include the *in situ*, synthetic and extracted satellite data from each class.

class may then correspond to conditions associated with matured to aged upwelling waters which usually consist of biomass ranges of 5-30 mg m^{-3} (Barlow, 1982; Barlow et al., 2005) where larger diatoms and various dinoflagellate species may occur depending on nutrient availability. Further investigation into the relationship between OWT dominance and upwelling is detailed in chapter 4.

While most of the moderate reflectance classes generally showed the sun-induced Chl *a* fluorescence peak at 681 nm, some of the spectra influencing the shapes of classes 5 and 6 displayed a 709 nm reflectance peak. This 709 nm peak is characteristic to high biomass phytoplankton dominated waters with [Chl *a*] greater than approximately 20 mg m^{-3} (Matthews et al., 2012), and occurs when the red peak shifts towards longer wavelengths as the 681 nm peak becomes masked by a combination of increased Chl *a* absorption, increased phytoplankton backscatter and pure water absorption. Moore et al. (2014) also found classes with characteristic peaks beyond 700 nm in their coastal/lake classification, particularly OWTs 4, 5 and 7 which had mean [Chl *a*] of 20.85, 206.37 and

36.00 mg m^{-3} respectively. Similarly, OWT6 in the current study represents the very high phytoplankton biomass or *bloom* class with mean [Chl *a*] of approximately 158 mg m^{-3}; this class most likely represents potentially harmful bloom forming dinoflagellates that often occur in the southern Benguela during the upwelling season. Although OWT6 does not specifically provide an indication of the presence of harmful toxins in the water, the use of this class can help to distinguish very high biomass blooms that could potentially lead to hypoxic events (Monteiro and van der Plas, 2006; Pitcher and Weeks, 2006).

Although distinguishing different phytoplankton biomass regions in the surface coastal waters around South Africa was one of the potential applications for using the OWT classification technique, [Chl *a*] on its own does not provide enough information to fully resolve water type characterization, since optically different waters may have overlapping [Chl *a*] (as seen previously in Moore et al., 2001, 2009, 2014) as demonstrated in figure 3.15. This effect is demonstrated in the current study when classes 3 and 4 showed similar [Chl *a*] ranges as some of the elevated reflectance classes, but represented different spectral shapes and underlying causalities. Classes 3 and 6 were the only OWTs in the moderate reflectance group where b_{bs} had a significant contribution to the dominant mode of variance, although its influence was still outweighed by the phytoplankton backscattering. Since there were no actual measurements of inorganic particulate matter or a_{dg} to inform the exact causes of the spectral variability, it is hard to tell with certainty whether the different shapes in R_{rs} presented in classes 3 and 6 were due to differences in water constituents and their IOPs, or rather due to the spectral variability introduced by the atmospheric correction of the satellite data.

The elevated reflectance classes are almost entirely made up of Case 2 synthetic data and MERIS data from the southern Benguela; the extracted satellite spectra included areas with high reflectance at 560 nm such as colour fronts in False bay (Shannon et al., 1991), and the turbid waters just off the Berg river in St Helena Bay or the Olifants river (approximately 100 km farther north) on the west coast. Although the dominant source of variance in the entire elevated reflectance database was b_{bs}, the phytoplankton IOPs and [Chl *a*] had a significant effect on the spectral shapes of the individual classes and were even the primary contributors to the dominant mode of variance of classes 7 and 11; OWT7 also had the highest mean [Chl *a*] of the elevated reflectance database, with mean values of 6.3 mg m^{-3}. Some of the spectra from OWT7 also expressed a 709 nm peak, suggesting high phytoplankton biomass conditions; these spectra could possibly represent a phytoplankton bloom in waters with additional sources of scattering such as

resuspended sediments or coccolithophores (Weeks et al., 2004). There was no significant correlation between the dominant causal optical variables and the reflectance of OWTs 9 and 10, which would suggest that these classes represent mixed water types with variable optical influences. Concentrations of phytoplankton may vary significantly independently of other water constituents in the elevated reflectance classes; appropriate and regionally calibrated Case 2 specific chlorophyll algorithms may be required to resolve the [Chl a] under these optically challenging conditions.

Whilst the moderate reflectance classes revealed clear ecologically-driven shifts in spectral shape, the increased optical complexity of the elevated reflectance classes made the causality of changes in spectral magnitude and shape less clear; the use of additional parameters, such as suspended particulate matter and particulate organic carbon (as used in Lubac and Loisel (2007)), may be needed to determine the ecological significance of these classes.

3.2.6 Class-specific Chl *a* algorithm selection

One of the primary objectives of the OWT classification of satellite imagery in this thesis is to facilitate application of water type appropriate Chl *a* algorithms; this ensures the highest quality in the returned Chl *a* products, preferably with estimates for associated errors and uncertainties. The following section presents the error and uncertainty analysis of the performance of five band ratio algorithms as tested on the *in situ* and synthetic R_{rs}. Once the best performing algorithm is selected per OWT, these algorithms are weighted and blended according to equations 2.20 and 2.21 and applied to the same *in situ* and synthetic R_{rs} to demonstrate the performance of the algorithm blending procedure.

Five different Chl *a* algorithms, which included the three blue-green (OCMe, OC3E, and OC4E) and two red-NIR (G2B and G3B) band ratio algorithms as detailed in section 2.3.4, were applied to solely the *in situ* data, as well as the combined *in situ* and synthetic data (represented as normal and bold font respectively in the results tables 3.3 - 3.6), from classes 1 to 11. Please note that a dedicated Case 2 algorithm was not assessed directly on the *in situ* or synthetic data. The performance of the standard MERIS Case 2 product, Algal2 (Doerffer and Schiller, 2007), was however assessed for the satellite matchup datasets in section 3.2.7; due to the nature of the neural network algorithm used to produce Algal2, it was not available for testing on the *in situ* or synthetic data.

The results from the various algorithms were compared to the corresponding known[1] [Chl *a*]. The scatter plots of the data from classes 1 to 6 (figure 3.16) and classes 7 to 11 (figure 3.18) complements the statistical analysis and facilitates appropriate algorithm choice by indicating areas ([Chl *a*]) of poor algorithm performance. The number of data-points (N) used for each calculation are given in table 3.3, whilst the RMSE (table 3.4), bias (table 3.5) and MARE (table 3.6), grouped together on pages 76 - 78, were used to discern the appropriate chlorophyll algorithm per class; the best indication was given by the RMSE, an inclusive metric that combines information of the mean and standard deviation in one term (Szeto et al., 2011). The bias provides indications of systematic offsets (Moore et al., 2014). Although the MARE and RMSE contain the term *error*, these statistics are used in the context of this thesis to denote the *uncertainty* of the modelled and satellite data, due to the discrete errors and uncertainty inherent to the *in situ* [Chl *a*] (Moore et al., 2009).

[1]The *known* [Chl *a*] can either be the measured *in situ* or synthetic [Chl *a*].

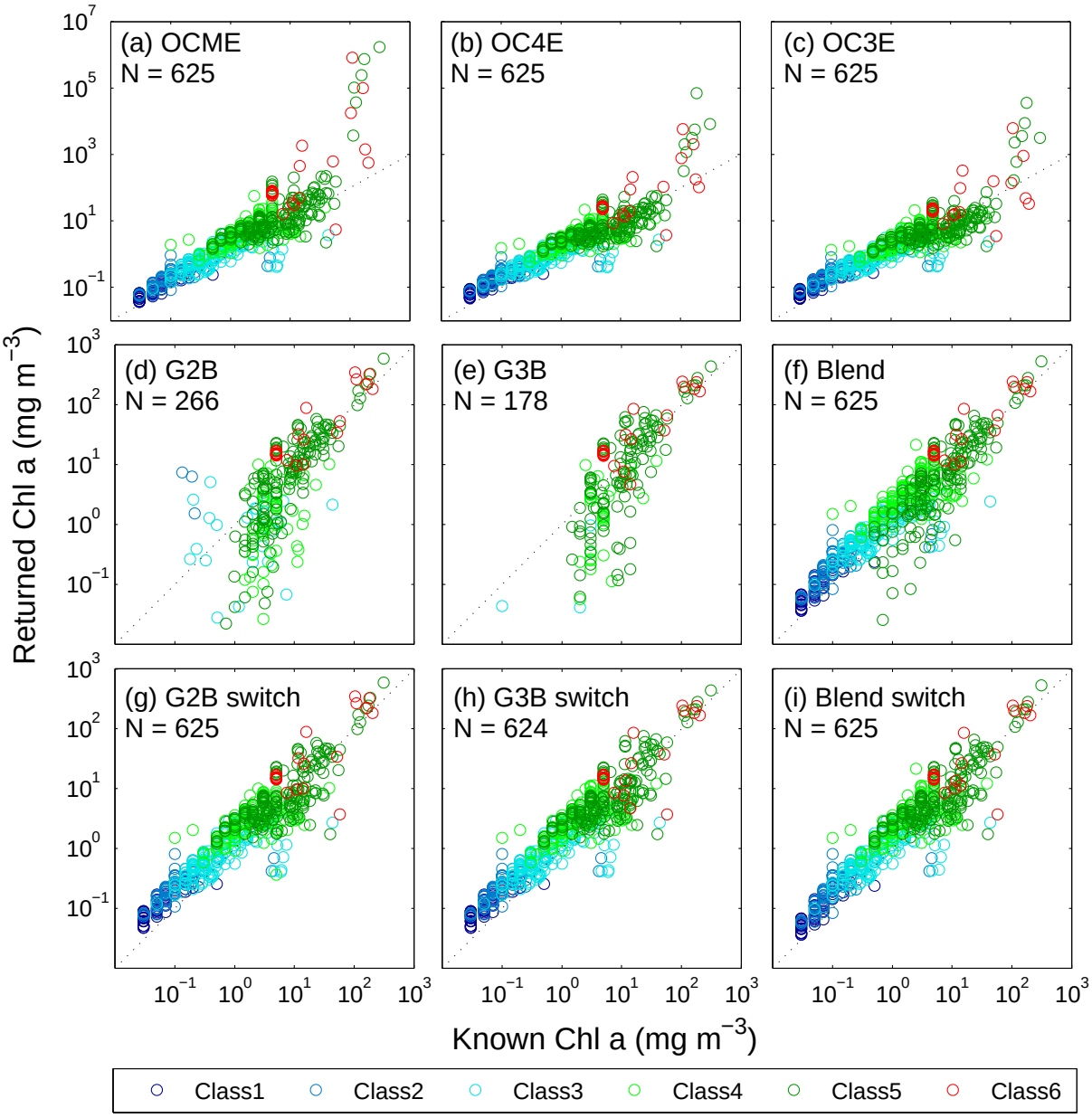


Figure 3.16: The modelled versus known (*in situ* and synthetic) chlorophyll *a* plots for classes 1 to 6. The modelled data represents the (a) OCMe, (b) OC4E, (c) OC3E, (d) G2B, (e) G3B and (f) blended products with different colours for each class. Panels (g), (h) and (i) show the results where an algorithm switch is implemented based on the values of OC4E; (g) represents the OC4E product with G2B results substituted >10 mg m^{-3}, whilst (h) represents the OC4E product with G3B results substituted >15 mg m^{-3}. The blended product (i) shows the results when the chosen algorithms are applied per OWT, as described at the end of this section. *N* is the number of data points with positive modelled [Chl *a*] for each plot. The dashed line represents the 1:1 line. (Note that the top row has a different y-axis scale)

From the scatter plots of the moderate reflectance classes (figure 3.16) it could be seen that the three blue-green band ratio algorithms OCMe (panel a), OC4E (panel b), and OC3E (panel c) performed well at low concentrations; all three models started to overestimate above approximately 10 mg m^{-3}, with OCMe having a greater overestimation than OC4E and OC3E. There also appeared to be some scatter above the 1:1 line around the known 0.1 mg m^{-3} [Chl *a*] step; this could potentially be an artefact of the Case 2 synthetic data which have optically significant constituents that do not covary with the phytoplankton IOPs. The red band-ratio algorithms performed better at higher [Chl *a*], with less scatter around the 1:1 line. Both the G2B (panel d) and G3B (panel e) algorithms show high amounts of scatter at concentrations close to and below 10 mg m^{-3}, as well as fewer data points below 10 mg m^{-3}; at [Chl *a*] <10 mg m^{-3} the R_{rs} signal in the NIR may be too low, which can result in larger measurement uncertainties (Dall'Olmo and Gitelson, 2006). The G3B algorithm appears to be more readily affected by the low signal, returning less available data at low *in situ* concentrations. In an attempt to minimize data loss due to negative returns from the red-NIR algorithms, a switching procedure was implemented based on the OC4E product; OC4E values would be returned where this blue-green algorithm was below a certain threshold, whilst either G2B or G3B would be implemented above the threshold. From the scatter plots and the literature (Dall'Olmo and Gitelson, 2006; Gilerson et al., 2010) 10 mg m^{-3} was used as this threshold for G2B algorithm, whilst a more conservative value of 15 mg m^{-3} was chosen for the G3B algorithm. These choices are supported by the algorithm comparison shown in figure 3.17, where the returns from the G2B and G3B algorithms are negative below OC4E values of approximately 10 and 15 mg m^{-3} for the data grouped into classes 5 and 6 respectively. There does not appear to be a visible step function or hard change between the algorithms in these plots, however it is possible that these switches may cause discontinuities within the blended [Chl *a*] satellite products; these effects will be assessed further in the next chapter. The results using the switch can also be seen in figure 3.16 panels g - i, and shows the use of more of the available data with less scatter around the 1:1 line.

The elevated reflectance classes (figure 3.18) were included separately to demonstrate the poor performance of the band ratio algorithms for these classes. This is particularly evident for classes 9 to 11 in the blue-green band ratio algorithms (panels a,b and c) at [Chl *a*]<2 mg m^{-3}; a potential reason for this may be the relatively strong and non-covarying contribution of b_{bs} and a_{dg}, which mask the phytoplankton signal and invalidate the empirical relationships that these algorithms are based upon. Although the blue-green band ratio algorithms appear to perform better for classes 7 and 8, they also show

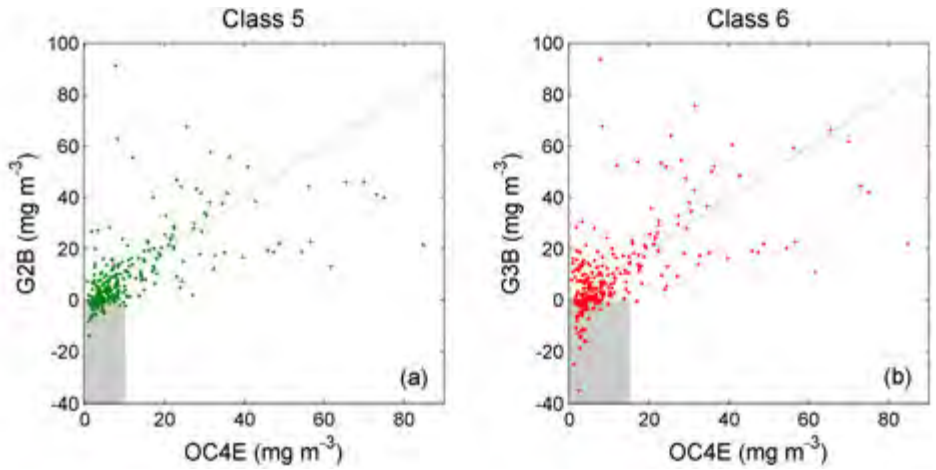


Figure 3.17: Algorithm performance of OC4E relative to the (a) G2B and (b) G3B algorithms for datapoints from classes 5 and 6 respectively. The shaded regions indicate conditions where red-NIR algorithms fail ([Chl *a*] <0) necessitating a transition between algorithms. The dashed line represents the 1:1 line.

high RMSEs and MAREs, in addition to a large positive bias (tables 3.4, 3.6, and 3.5 respectively). The red-NIR algorithms show similarly poor performance, which could potentially be attributed to the low signal in the red associated with [Chl *a*]<10 mg m^{-3}; this signal may also be masked by other optically significant constituents. In the tables the results for classes 7 to 11 (italicized) are only shown for completeness; only the spectra from the moderate reflectance classes were used for the sum total calculations and the weighting and blending of the data, since none of the band-ratio algorithms displayed adequate performance in the high reflectance classes. Other algorithm types that are optimized for highly scattering waters are thus required for these classes, which will be demonstrated in the next section.

Overall the RMSE value (table 3.4) was used to choose the optimum algorithm per class; generally both the RMSE for the *in situ* and the combined data would be lowest for the chosen class. The only class where this was not the case was class 5, where OC3E showed the lowest RMSE; however, class 5 contained spectra representing [Chl *a*] of up to 309 mg m^{-3} with approximately 29 % of the values >10 mg m^{-3}, a range where the OC3E had shown poor performance. In this case it would be more appropriate to switch to a red-NIR algorithm (G2B in the case of class 5) to resolve these high [Chl *a*]. OC4E was applied as the default product, instead of OC3E, as it showed the best performance in the switching scenario, decreasing the RMSE for class 5 from 0.475 to 0.337 (not shown in tables). A similar switch (at 15 mg m^{-3}) was applied for class 6 using G3B and OC4E, which changed RMSE results for class 6 from 0.403 to 0.462 (not shown in tables); although the switch

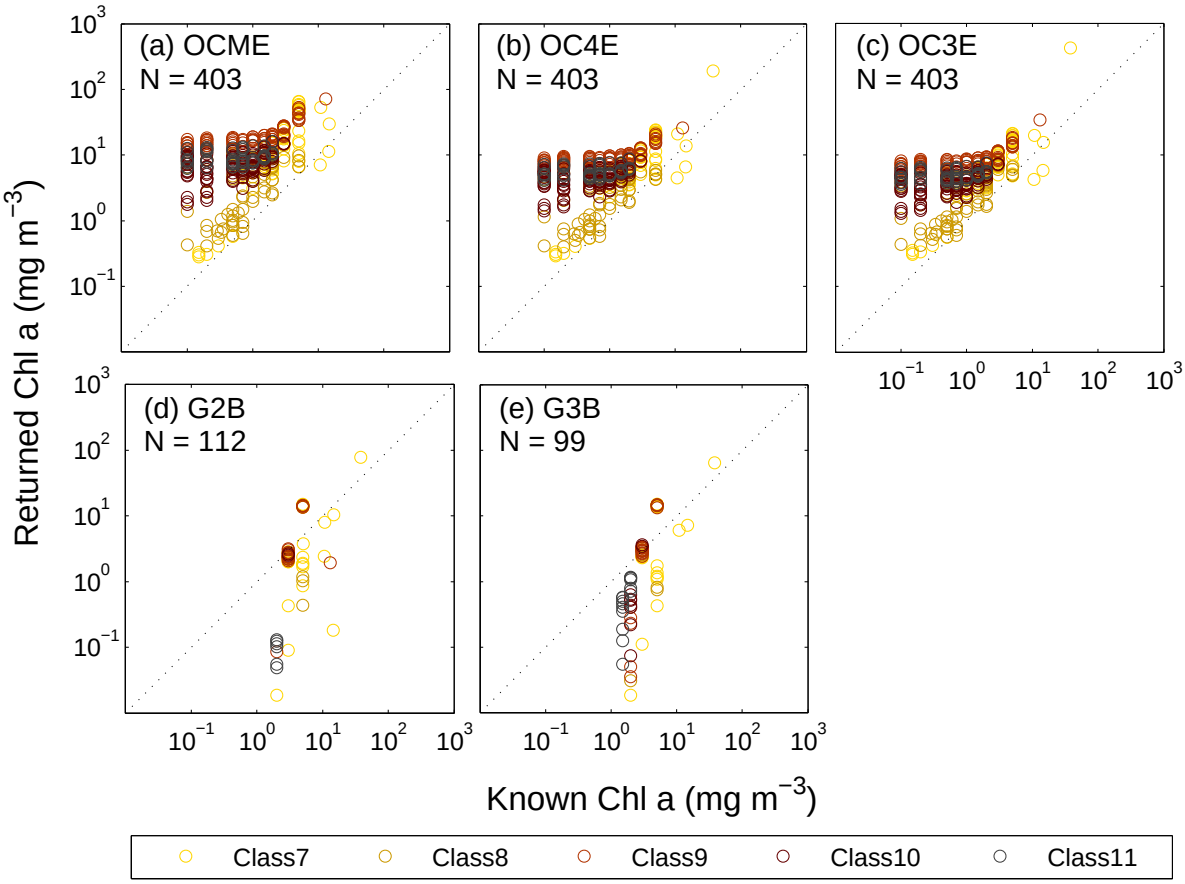


Figure 3.18: The modelled versus known (*in situ* and synthetic) chlorophyll *a* plots for classes 7 to 11. The modelled data represents the (a) OCMe, (b) OC4E, (c) OC3E, (d) G2B, and (e) G3B products with different colours for each class. N is the number of data points per plot. The dashed line represents the 1:1 line.

increased the RMSE of class 6 in the *in situ* data, it greatly improved the RMSE for the satellite matchup data (shown in table 3.7 in section 3.2.7). When applying both switching algorithms during the blending procedure, the RMSE for all the data decreased from 0.363 to 0.354 (not shown in tables). Although the red-NIR algorithms appeared to give the best performance for the elevated reflectance classes (based on the RMSE), these algorithms had a tendency to produce predominantly negative [Chl *a*], resulting in many data being discarded from the analyses; these algorithms would thus not necessarily be adequate algorithms for operational use with classes 7 to 11. The MARE was primarily lowest for the RMSE-chosen class. The scatter plots and the bias results indicate that the blue-green algorithms have a tendency to overestimate, particularly for the higher biomass and elevated reflectance classes, whilst the red-NIR algorithms show a tendency to underestimate at the lower concentration ranges.

Ultimately the algorithms used in the blending procedure were: OCMe for classes 1 and 2, OC4E for classes 3 and 4, as well as OC4E for classes 5 and 6 (which would switch to G2B when reaching [Chl *a*] of >10 mg m^{-3} for class 5, and to G3B when reaching [Chl *a*] of >15 mg m^{-3} for class 6). During the clustering process a spectrum is grouped into the clusters to which it shows maximum membership, even though the spectrum could have had partial membership to some of the other clusters. The blending method applies the classification algorithm to the R_{rs} and uses the membership functions as weights for the class-specific chlorophyll algorithm application and blending procedure. The blended product gave either very similar or improved RMSE, bias and MARE results compared to the best performing algorithm for each class. The results were particularly improved for the blended product when using all the *in situ* and synthetic data, where the blended product showed the lowest bias (0.128) and RMSE (0.363). Although the MARE for the blended product (69.6 %) was slightly higher than for the G2B and G3B, all data were used in the calculation for the blend, whilst only 28 to 42 % of the data could be used for the red-NIR algorithms. The choice of algorithms are supported by the scatter plot of the blended product (figure 3.16f), whilst the use of the switch for classes 5 and 6 produces the least amount of scatter around the 1:1 line (figure 3.16i); this demonstrates the optimal integration and performance of switching and blended Chl *a* algorithms.

Table 3.3: The number of datapoints N used for the bias and RMSE calculations per class and per algorithm. For each algorithm the results when using only the *in situ* data (tsrb) are displayed first, whilst the second column shows the results when using all *in situ* and synthetic data. The modelled data included OCMe, OC4E, OC3E, G2B and G3B Chl *a* products. The blended product (in bold font) shows the result when using the best performing algorithm (based on the RMSE and indicated by an asterisk) for each of the first six classes to appropriately weight and blend the Chl *a* algorithms according to equations 2.20 and 2.21. Data and results from classes 7 - 9 (italicized) are shown for completeness, but were not included in the algorithm calculations using all the data (All).

Class	OCMe (tsrb)	**OCMe**	OC4E (tsrb)	**OC4E**	OC3E (tsrb)	**OC3E**	G2B (tsrb)	**G2B**^{n}	G3B (tsrb)	**G3B**^{n}	Blend (tsrb)	**Blend**
1	0	**58***	0	**58**	0	**58**	0	**0**	0	**0**	0	**58**
2	8	**78***	8	**78**	8	**78**	3	**3**	0	**0**	8	**78**
3	74	**141**	74	**141***	74	**141**	22	**25**	0	**4**	73	**140**
4	43	**145**	43	**145***	43	**145**	26	**59**	10	**34**	43	**145**
5	122	**179**	122	**179**	122	**179**	115	**155***	83	**116**	122	**179**
6	13	**24**	13	**24**	13	**24**	13	**24**	13	**24***	13	**24**
7	*6*	***69***	*6*	***69***	*6*	***69***	*6*	***41***	*4*	***38***	*0*	***0***
8	*6*	***83***	*6*	***83***	*6*	***83***	*0*	***6***	*0*	***6***	*0*	***0***
9	*1*	***88***	*1*	***88***	*1*	***88***	*1*	***27***	*0*	***30***	*0*	***0***
10	*0*	***105***	*0*	***105***	*0*	***105***	*0*	***2***	*0*	***9***	*0*	***0***
11	*0*	***58***	*0*	***58***	*0*	***58***	*0*	***6***	*0*	***16***	*0*	***0***
Total	260	**625**	260	**625**	260	**625**	179	**266**	106	**178**	259	**624**

^{n} excludes negative data

* best performing [Chl *a*] algorithm per class (based on RMSE)

Table 3.4: The root mean squared error (RMSE) in log10 units per class and per algorithm. For each algorithm the results for solely the *in situ* data (tsrb) are displayed first, whilst the second column shows the results when using all *in situ* and synthetic data (in bold font). The modelled data included OCMe, OC4E, OC3E, G2B and G3B Chl *a* products. The blended product shows the result when using the best performing algorithm (based on the RMSE and indicated by an asterisk) for each of the first six classes to appropriately weight and blend the Chl *a* algorithms according to equations 2.20 and 2.21. Data and results from classes 7 - 11 (italicized) are shown for completeness, but were not included in the sum total (All) algorithm calculations.

Class	OCMe (tsrb)	**OCMe**	OC4E (tsrb)	**OC4E**	OC3E (tsrb)	**OC3E**	G2B (tsrb)	**G2B^{n}**	G3B (tsrb)	**G3B^{n}**	Blend (tsrb)	**Blend**
1	-	**0.248***	-	**0.335**	-	**0.336**	-	-	-	-	-	**0.248**
2	0.453	**0.373***	0.469	**0.404**	0.465	**0.416**	1.421	**1.421**	-	-	0.464	**0.372**
3	0.375	**0.376**	0.362	**0.351***	0.368	**0.357**	0.861	**0.837**	-	**1.014**	0.340	**0.337**
4	0.420	**0.527**	0.378	**0.397***	0.397	**0.396**	0.835	**0.785**	0.705	**0.718**	0.374	**0.393**
5	0.940	**0.862**	0.460	**0.452**	0.461	**0.444**	0.449	**0.475***	0.548	**0.567**	0.355	**0.381**
6	1.719	**1.485**	0.809	**0.763**	0.823	**0.740**	0.334	**0.411**	0.318	**0.403***	0.307	**0.392**
7	*0.745*	***0.772***	*0.373*	***0.493***	*0.501*	***0.462***	*0.835*	***0.628***	*0.386*	***0.580***	-	**-**
8	*0.394*	***0.784***	*0.334*	***0.641***	*0.330*	***0.592***	-	***0.581***	-	***0.874***	-	**-**
9	-	***1.360***	-	***1.101***	-	***1.051***	-	***0.428***	-	***0.578***	-	**-**
10	-	***1.199***	-	***1.019***	-	***0.953***	-	***0.066***	-	***0.732***	-	**-**
11	-	***1.266***	-	***1.062***	-	***1.011***	-	***1.362***	-	***0.627***	-	**-**
All	0.799	**0.646**	0.447	**0.420**	0.453	**0.418**	0.603	**0.610**	0.543	**0.594**	0.356	**0.363**

^{n} excludes negative data

* best performing [Chl *a*] algorithm per class (based on RMSE)

Table 3.5: Bias in log10 units per class and per algorithm. For each algorithm the results for solely the *in situ* data (tsrb) are displayed first, whilst the second column shows the results when using all *in situ* and synthetic data (in bold font). The modelled data included OCMe, OC4E, OC3E, G2B and G3B Chl *a* products. The blended product shows the result when using the best performing algorithm (based on the RMSE and indicated by an asterisk) for each of the first six classes to appropriately weight and blend the Chl *a* algorithms according to equations 2.20 and 2.21. Data and results from classes 7 - 11 (italicized) are shown for completeness, but were not included in the sum total (All) algorithm calculations.

Class	OCMe (tsrb)	**OCMe**	OC4E (tsrb)	**OC4E**	OC3E (tsrb)	**OC3E**	G2B (tsrb)	**G2B^{n}**	G3B (tsrb)	**G3B^{n}**	Blend (tsrb)	**Blend**
1	-	**0.210***	-	**0.306**	-	**0.308**	-	**-**	-	**-**	-	**0.210**
2	-0.151	**0.281***	-0.147	**0.319**	-0.129	**0.335**	1.370	**1.370**	-	**-**	-0.159	**0.279**
3	0.038	**0.183**	-0.036	**0.124***	-0.041	**0.123**	-0.357	**-0.382**	-	**-0.870**	-0.030	**0.111**
4	0.125	**0.404**	-0.060	**0.231***	-0.084	**0.216**	-0.586	**-0.559**	-0.321	**-0.469**	-0.071	**0.218**
5	0.367	**0.444**	-0.002	**0.123**	-0.078	**0.063**	-0.102	**-0.118***	-0.148	**-0.164**	-0.065	**-0.047**
6	1.205	**1.178**	0.375	**0.526**	0.238	**0.416**	0.195	**0.329**	0.122	**0.287***	0.139	**0.312**
7	*0.444*	***0.701***	*0.057*	***0.434***	*0.115*	***0.405***	*-0.441*	***-0.081***	*-0.236*	***-0.028***	-	**-**
8	*0.386*	***0.658***	*0.326*	***0.506***	*0.323*	***0.462***	-	***-0.452***	-	***-0.583***	-	**-**
9	-	***1.284***	-	***0.995***	-	***0.940***	-	***0.062***	-	***-0.029***	-	**-**
10	-	***1.130***	-	***0.938***	-	***0.865***	-	***-0.065***	-	***-0.584***	-	**-**
11	-	***1.197***	-	***0.979***	-	***0.924***	-	***-1.351***	-	***-0.533***	-	**-**
All	0.259	**0.362**	-0.007	**0.205**	-0.054	**0.182**	-0.157	**-0.183**	-0.131	**-0.177**	-0.049	**0.128**

^{n} excludes negative data

* best performing [Chl *a*] algorithm per class (based on RMSE)

Table 3.6: The median absolute relative error (MARE) per class and per algorithm. For each algorithm the results for solely the *in situ* data (tsrb) are displayed first, whilst the second column shows the results when using all *in situ* and synthetic data (in bold font). The modelled data included OCMe, OC4E, OC3E, G2B and G3B Chl *a* products. The blended product shows the result when using the best performing algorithm (based on the RMSE and indicated by an asterisk) for each of the first six classes to appropriately weight and blend the Chl *a* algorithms according to equations 2.20 and 2.21. Data and results from classes 7 - 11 (italicized) are shown for completeness, but were not included in the sum total (All) algorithm calculations.

Class	OCMe (tsrb)	**OCMe**	OC4E (tsrb)	**OC4E**	OC3E (tsrb)	**OC3E**	G2B (tsrb)	**G2B^{n}**	G3B (tsrb)	**G3B^{n}**	Blend (tsrb)	**Blend**
1	-	**69.1***	-	**104**	-	**104**	-	**5851**	-	**2222**	-	**69.1**
2	34.9	**105***	37.9	**121**	44.2	**131**	233	**2739**	28753	**1353**	30.8	**103**
3	31.7	**86.4**	31.7	**70.7***	33.0	**73.2**	212	**372**	1813	**651**	35.0	**60.1**
4	59.2	**186**	55.4	**96.1***	61.9	**102**	97.9	**129**	259	**186**	54.7	**94.4**
5	73.4	**123**	36.2	**58.2**	43.7	**59.3**	41.7	**60.4***	76.1	**92.9**	41.5	**47.3**
6	692	**1272**	93.6	**399**	83.9	**292**	58.7	**184**	34.4	**176***	34.4	**176**
7	*108*	***442***	*55.8*	***224***	*59.7*	***188***	*53.3*	***138***	*73.7*	***168***	-	**-**
8	*131*	***271***	*104*	***161***	*108*	***127***	*702*	***352***	*19504*	***326***	-	**-**
9	*451*	***1198***	*97.3*	***576***	*161*	***504***	*85.3*	***189***	*223*	***194***	-	**-**
10	-	***906***	-	***527***	-	***423***	-	***484***	-	***355***	-	**-**
11	-	***1097***	-	***641***	-	***556***	-	***339***	-	***182***	-	**-**
All	65.1	**114**	38.4	**89.5**	43.8	**84.8**	89.8	**196**	194	**286**	38.8	**71.2**

^{n} does not exclude negative data

* best performing [Chl *a*] algorithm per class (based on RMSE)

3.2.7 Validation of blended Chl *a* algorithm performance using matchup MERIS data

In order to characterize the MERIS-specific errors and uncertainties associated with the Chl *a* algorithms, each of the five Chl *a* algorithms utilized in section 3.2.6 were also applied to MERIS matchups. The matchups consisted of two different validation datasets, in addition to a combination of the two; the three datasets are characterized as 48 datapoints from the southern Benguela (mean [Chl *a*] 26.8 ± 55.5 mg m^{-3}), 241 from the MERMAID database (mean [Chl *a*] 2.14 ± 3.23 mg m^{-3}), and 289 for the combination of the two (mean [Chl *a*] 6.23 ± 24.4 mg m^{-3}). The Algal2 product, a standard MERIS Case 2 chlorophyll product produced by a neural network algorithm (Doerffer and Schiller, 2007), is also utilized in this section; due to the very poor performance demonstrated by all the band ratio algorithms when spectra represented the moderate to highly scattering waters of the elevated reflectance classes, the Algal2 product was used in the algorithm blending procedure when spectra were classified as OWTs 7 to 11. In an attempt to minimize data loss due to negative returns from the G2B and G3B algorithms, a procedure was implemented which returned results from the available blue-green ratio algorithms as default, but would switch to the red-NIR algorithms above a certain threshold; following discussions from section 3.2.6, OC4E was chosen as the default product, whilst G2B was applied >10 mg m^{-3}, and G3B was applied >15 mg m^{-3} [Chl *a*]. The statistical results for the three matchup datasets can be seen in tables 3.7 to 3.9.

When applied individually, the OC4E and G2B gave the lowest RMSE and MARE results out of the standard algorithms for the MERMAID and Benguela datasets respectively. In all cases the switch facilitates the utilization of more of the available satellite data for both G2B and G3B, as seen in table 3.7; however, the MERMAID and combined datasets contained two elevated reflectance spectra that did not conform to any of these OWT classes, and were thus unclassified and not used for the creation of the blended product. In the southern Benguela the blended product (without any switch) gives the lowest RMSE and MARE, whilst the switch produces a similar but slightly higher values for the blend; this may be because the average [Chl *a*] of the dataset is relatively high and the G2B was already the best performing algorithm for the dataset. However, there is a clear decrease in the RMSE and MARE results of the G2B, G3B and blended product when the switch is implemented solely for the MERMAID database. One could thus argue that the use of the switch is more prudent for areas which may include both low and high [Chl *a*]. When the algorithm switch is incorporated into the algorithm blend it

also decreased the MARE of the MERMAID database blended product by approximately 43% (table 3.8).

In all cases either the G2B or G3B switching algorithms gave similar or slightly lower RMSE than the blended product; these results highlight the effectiveness of a simple switching algorithm. However, the switching algorithms do not have the inherent capacity to automatically account for and resolve the potential simultaneous occurrence of high b_{bs} and low biomass, as done within the OWT classification and blending framework. Upon closer inspection the slightly elevated RMSE values for the blended product resulted from the inclusion of the Algal2 algorithm; there is thus a clear need for better performing Case 2 algorithms, and it is very likely that future algorithms that incorporate regional IOP parameterizations could decrease the RMSE for the blended product.

The combined use of the blended product with the switching threshold consistently provided lower RMSE results than any other single algorithm, with the lowest uncertainties (45.5% across the blended product for the combined dataset) and a slightly positive bias (0.087 across the blended product for the combined dataset). This method automatically applies the optimal algorithms per reflectance spectrum, and can operate successfully across a wide dynamic range of [Chl a] spanning almost four orders of magnitude, as well as across a wide range of OWTs. Although scattering waters are not specifically dealt with and resolved in as much detail as the phytoplankton dominated water classes, they still provide the capability to recognize regions within satellite images where larger uncertainties may occur within the current algorithm framework (i.e. moderate to highly scattering waters).

Table 3.7: RMSE for the OCMe, OC4E, OC3E, Algal2, G2B, G3B and blended chlorophyll *a* products (in log10 units) for extracted MERIS matchup data from the southern Benguela (SB) and the MERMAID database (MM). The "switch" indicates that both the OC4E-G2B and OC4E-G3B algorithm switches, as described in section 3.2.6, were implemented. The values in the square brackets represent the amount of datapoints that were used to calculate the statistics for that specific product. The asterisk represents the lowest RMSE per dataset.

Dataset	OCMe	OC4E	OC3E	Algal2	G2B	G3B	Blend
SB	1.093	0.604	0.974	0.488	0.480 [N=40]	0.652 [N=46]	0.399* [N=48]
SB switch	1.093	0.604	0.974	0.488	0.399 [N=47]	0.394* [N=47]	0.428 [N=48]
MM	0.503	0.377*	0.422	0.440	0.915 [N=70]	1.518 [N=127]	0.701 [N=239]
MM switch	0.503	0.377	0.422	0.440	0.340* [N=240]	0.342 [N=241]	0.369 [N=239]
ALL switch	0.640	0.423	0.553	0.448	0.350* [N=287]	0.351 [N=288]	0.380 [N=287]

Table 3.8: The median absolute relative error (MARE) for the OCMe, OC4E, OC3E, Algal2, G2B, G3B and blended chlorophyll *a* products for extracted MERIS matchup data from the southern Benguela (SB) and the MERMAID database (MM). The "switch" indicates that both the OC4E-G2B and OC4E-G3B algorithm switches, as described in section 3.2.6, were implemented. The asterisk represents the lowest MARE per dataset.

Dataset	OCMe	OC4E	OC3E	Algal2	G2B	G3B	Blend
SB	89.2	52.7	56.3	63.8	48.0	49.3	44.4*
SB switch	89.2	52.7	56.3	63.8	47.2	46.2*	47.2
MM	67.4	45.8*	50.5	51.5	523	1673	88.6
MM switch	67.4	45.8	50.5	51.5	44.8*	45.8	45.0
ALL switch	71.0	46.9	52.3	52.0	45.7	45.8	45.5*

Table 3.9: Bias for the OCMe, OC4E, OC3E, Algal2, G2B, G3B and blended chlorophyll *a* products (in log10 units) for extracted MERIS matchup data from the southern Benguela (SB) and the MERMAID database (MM). The "switch" indicates that both the OC4E-G2B and OC4E-G3B algorithm switches, as described in section 3.2.6, were implemented. The asterisk represents the lowest bias per dataset.

Dataset	OCMe	OC4E	OC3E	Algal2	G2B	G3B	Blend
SB	0.392	0.024	0.075	-0.153	0.004*	0.195	-0.050
SB switch	0.392	0.024	0.075	-0.153	-0.001*	-0.020	-0.026
MM	0.240	0.131	0.117	-0.025*	0.170	0.974	-0.117
MM switch	0.240	0.131	0.117	-0.025*	0.112	0.120	0.101
ALL switch	0.266	0.113	0.110	-0.046*	0.093	0.098	0.087

3.2.8 The potential for weighted uncertainty maps for blended Chlorophyll products

The ocean colour user community has identified a need for per pixel uncertainty estimates for satellite products. Consequently Moore et al. (2015) and Mélin et al. (2016) have derived uncertainty estimates for R_{rs} based on global *in situ*/satellite matchup sets and coincident daily records from different sensors, respectively; the Ocean Colour Climate Change Initiative (OC-CCI) programme has also generated a validated and error-characterized set of essential climate change variables from satellite observations for Case 1 waters (Grant et al., 2015), whilst Moore et al. (2009) have characterized the Chl *a* uncertainty for MODIS Aqua OC3M associated with different OWTs.

The current study offers a preliminary demonstration of blending the uncertainties from the different Chl *a* algorithms based on the performance per OWT. The maps uses the MARE as derived per OWT from the combined southern Benguela and MERMAID dataset; the MARE value that is used in the uncertainty weighting procedure is assigned according to the algorithms used to create the blended [Chl *a*] product. The blended uncertainty product was created using equations 2.22 and 2.23; based on the MARE results in table B.1 in appendix B, the error assigned to each OWT were as follows:

OWT1 = 21.9%	**OWT2 = 37.8%**	**OWT3 =45.4%**	**OWT4 = 73.9%**
OWT5 = 28.1%	**OWT6 = 70.0%**	**OWT7 = 78.7%**	**OWT8 = 78.1%**
OWT9 = [78.0%]	**OWT10 = 37.7%**	**OWT11 = 47.2%**	

Note that the MARE for OWT9 could not be calculated since no matchup data were assigned to this water type; for the purpose of satellite image application a tentative estimate of 78% was used (based on the values of OWTs 7 and 8).

This method produces blended uncertainty maps that demonstrate the performance of the different Chl *a* algorithms for the MERIS sensor. It should be noted that these uncertainty estimates only present a preliminary demonstration of the potential capability, and should not be taken to represent the optimal values for South African coastal waters, particularly due to the low amount of data within some of the OWTs (as seen in table B.2 in appendix B). Improved error and uncertainty estimates can be determined once more *in situ*/satellite matchup data become available, particularly from next generation sensors such as the ocean and land colour instrument (OLCI) on board Sentinel-3, which will have spectral coverage in the red-NIR desirable for resolving water constituents in

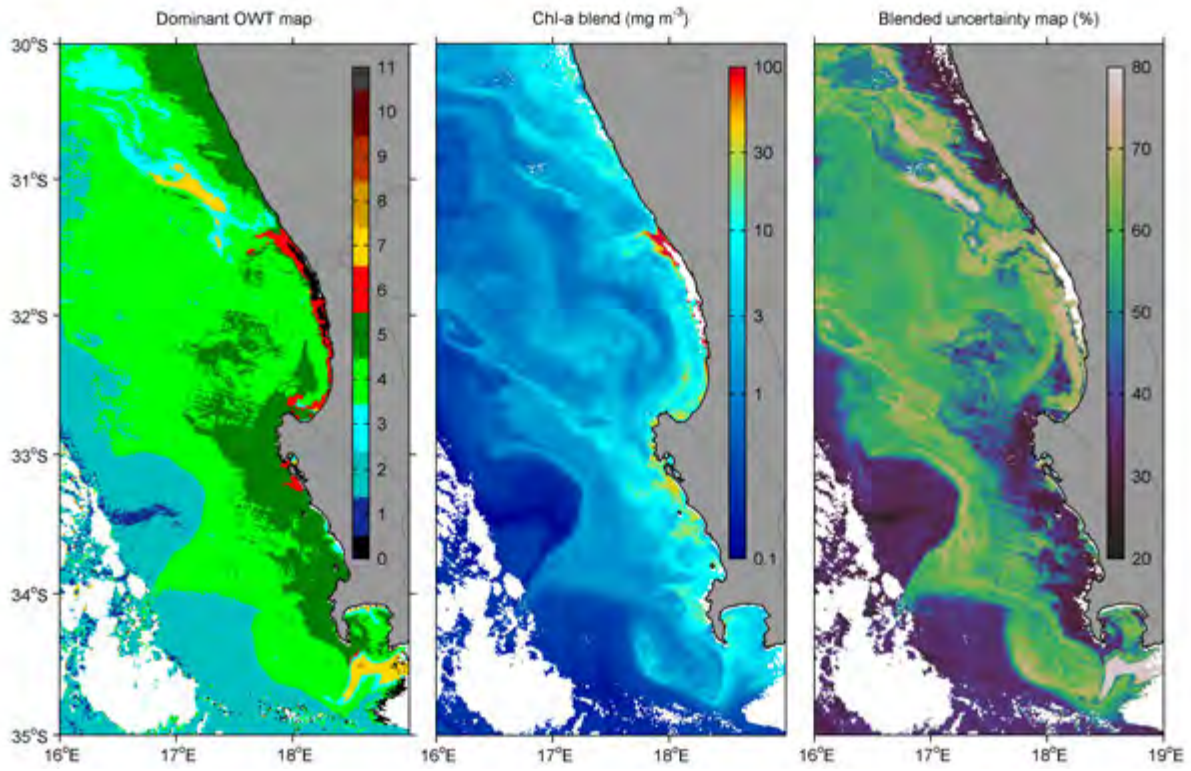


Figure 3.19: The dominant OWT, blended [Chl *a*] product and blended uncertainty product for the 14th of March 2009 over the southern Benguela. White areas represent clouds; in the blended [Chl *a*] and uncertainty maps white areas also represents pixels that were flagged for invalid reflectance.

eutrophic and highly turbid waters. An example is provided in figure 3.19 which shows the blended uncertainty product that accompanies the blended [Chl *a*] product map.

3.3 Discussion of results

3.3.1 Relevance and soundness of clustering and classification methods driven by application

The choice of the clustering approach is dependent on the application for which it is envisaged (Mélin and Vantrepotte, 2015). For the current study the primary focus was to develop [Chl *a*] products that were optimized across many water types; this required the characterization of the OWTs present in South African coastal waters, whilst maintaining a strong focus on phytoplankton biomass. Secondary applications included the informative characterization of the bio-optical and/or ecological provinces in a satellite image directly from R_{rs} without the need for a derived Chl *a* product, as well as the direct identification of specific key water types as stand-alone products (e.g. very high biomass dinoflagellate blooms). The OWTs and classification framework utilized in the current body of work differs from many previous studies (e.g. Vantrepotte et al., 2012; Mélin and Vantrepotte, 2015) in that it includes wavelengths in the red-NIR; the importance of this spectral region has been demonstrated for OWT classification (Moore et al., 2014) as well as the estimation of [Chl *a*] in turbid productive waters (Gilerson et al., 2010; Moses et al., 2012) and the southern Benguela (Bernard et al., 2005; Matthews et al., 2012). With the availability of good spectral coverage in the red-NIR the Level 2 radiometric data from MERIS and the classification methods as applied in Moore et al. (2001, 2014) provided the ideal data and framework to resolve [Chl *a*] over a wide range of environmental conditions.

In terms of creating the OWTs there is a plethora of algorithms available for dataset clustering, each of which may result in slightly different clusters; even with the FCM algorithm the results may vary depending on the initial starting conditions that are used. Moore et al. (2014) noted that the k-means algorithm gave similar clustering results and could also have been used for their study. The Agglomerative Ward's Linkage (AWL) (Ward, 1963) and k-means (Hartigan, 1975) were tested on the combined dataset (dataset 6 in figure 2.2) and the resulting clusters (not shown) were very similar to those from steps six and seven; however, as with the FCM algorithm, these algorithms also did not discriminate the different biomass groupings at low reflectance. To a certain extent the clustering algorithm or method is of lesser importance, provided that it produces meaningful spectral groupings and well understood classification results with respect to its intended use. For the purpose of the current study it was vital that the clustering be optimized according to carefully chosen ecological criteria.

A known limitation of the FCM clustering method is the unknown variable of the appropriate amount of clusters to use; although validity functions exist that can help with this decision (Windham, 1982; Xie and Beni, 1991), there is no one method that is completely reliable (Moore et al., 2001). In this study four validity functions were used to provide quantitative evaluations of the cluster structure imposed on the data. Although these functions were used to guide the clustering process, they do not provide information about the application or ecological relevance of the clusters. Thus clustering is subjective and the applications conditional upon the region of interest. Much of the cluster related decision-making relies on visual inspection and inter-comparison of the cluster structures on classified test images, whilst assessing the cluster-classification performance with regard to the desired applications.

The Mahalanobis distance is a very popular method for classification of satellite imagery ocean colour reflectance (Fiscella et al., 2000; Moore et al., 2001; Alimonte et al., 2003; Moore et al., 2009; Mélin et al., 2011; Vantrepotte et al., 2012; Moore et al., 2012, 2014; Mélin and Vantrepotte, 2015) and was adopted for image classification in this study. The statistics that this image classification technique relies on are the mean and the covariance matrix of each cluster/OWT; thus for successful image classification it is vital that these statistics are representative of the relevant range of regional water types. The Mahalanobis distance is affected by the choice of wavelengths and their accuracies used in the classification (Moore et al., 2014), while its performance may also be affected by the sample size of the clusters, since the covariance matrix of a small cluster may vary significantly from the true population covariance matrix (Hoffbeck and Landgrebe, 1996; Moore et al., 2001).

The use of reflectance normalization before clustering is also a popular approach (Lubac and Loisel, 2007; Vantrepotte et al., 2012; Mélin and Vantrepotte, 2015). This process removes some of the first order variability associated with spectral amplitude changes due to backscattering of non-algal particulate matter, whilst driving the focus on the spectral shape of the data. It has been suggested by Mélin and Vantrepotte (2015) that normalization of the R_{rs} allows a more dynamic range of constituent concentrations per class which may be beneficial for algorithm development, but that this approach would not be as useful for applications aimed at detecting thresholds (Mélin and Vantrepotte, 2015); hence for the purposes of this study, where OWT classification was partly aimed at distinguishing different levels of biomass, it is more desirable to use unnormalized R_{rs} spectra.

3.3.2 The building blocks of a clustering database: the effects of database composition

In situ data offer a desirable starting point for a training[1]dataset in the clustering process since these data are considered to have smaller uncertainties compared to satellite data and often includes coincident ancillary data (Mélin and Vantrepotte, 2015) that could be used for cluster characterisation and algorithm development. However, many regions may not be sampled due to the monetary, time and spatial constraints associated with data collection, e.g., in the current study where the data collection efforts were focussed almost entirely on the southern Benguela, with very few data collected along the south and east coast of South Africa. Even where data are collected it often does not cover all the optical variability that may be found within a certain region. Clustering training sets consisting of only *in situ* data are thus constrained by the ranges of these data, limiting the extent to which the OWTs would be representative of different seasons or other regions (Mélin and Vantrepotte, 2015). This sentiment is echoed in this study where dataset 1, consisting of only *in situ* data, showed very low relative clustering success compared to other datasets (as seen in table 3.1 in section 3.2.1); although low clustering success could be an artefact of small cluster size, it is also very likely the result of uncaptured OWTs.

Previous studies have made use of global *in situ* or satellite R_{rs} data to create comprehensive cluster training data sets. Global *in situ* data sets, such as the NOMAD data set (Werdell and Bailey, 2005), have been used for global clustering and classification studies (Moore et al., 2009), whilst Mélin and Vantrepotte (2015) extracted satellite data for the coastal domain accounting for 12% of the Earth's surface to use as a training set for clustering. Although these datasets may account for many of the possible OWTs that can be found all over the world, it is improbable that these datasets can cover all possible R_{rs} eventualities, particularly highly productive and optically complex waters. In order to compare the OWTs from the current study to globally defined OWTs, the coastal OWTs defined in Moore et al. (2001, 2009) were applied to the same test images used previously to define the classification success (table 3.1) for all three domains; the Moore coastal OWTs have been made available through the Basic ENVISAT Toolbox for (A)ATSR and MERIS (BEAM) with the *OWT Classification processor* plug-in (v 1.7). The Moore OWT classification resulted in $\bar{A}$ values of 0.393 ± 0.273, 0.457 ± 0.219, and 0.552 ±

[1]A dataset that is used as input during the data clustering process is often referred to as a *training dataset* since it is used to *train* the clusters that would subsequently be utilized during the classification process.

0.206 for the west, south and east coast respectively. When compared to the $\bar{A}$ shown in table 3.1, the OWTs from the current study performed better within the west and south coast domains (0.783 and 0.677 respectively), whilst the Moore coastal OWTs showed higher classification success for the east coast domain. The bio-optical conditions along the east coast could possibly conform more closely to global coastal conditions, whilst the phytoplankton-dominated productive waters of the west and south coasts could benefit from more specialized OWT classes and sensor optimization, as presented in the current study.

Both the Moore et al. (2009) and Mélin and Vantrepotte (2015) studies were limited to the spectral resolution between 412 and 670 nm which constrained the ability to differentiate certain optical signatures of turbid and/or productive estuarine and coastal waters that may be found in the red and NIR spectral regions. Moore et al. (2014) partly addressed this issue when they clustered hyperspectral *in situ* R_{rs} from lakes and coastal waters which included spectral data up to 750 nm; as a result they could distinguish a variety of new OWTs in the productive and optically complex waters of Lake Erie. The OWTs from Moore et al. (2014) are also available through the BEAM *OWT Classification processor*, and were used to classify the test images for the productive west coast domain, but only resulted in relatively low $\bar{A}$ of 0.205 ± 0.163; these results further supports the value of specialized or regionally-optimized OWTs within this domain.

The effects of adding synthetic data

Synthetic data rely on validated and well understood IOP and radiative transfer models to derive the R_{rs} and can provide useful additional data for cluster characterisation and algorithm development - if appropriately synthesized and constrained. To some extent the addition of synthetic data is similar to the use of global data for a regional study: although the data represent plausible R_{rs} spectra, they may not be representative of the waters found in the study region. In such a case the onus falls on the regional cluster developer to apply regional knowledge to ensure that the resulting clusters produce physically meaningful classification results and, when necessary, to exclude data that are not characteristic of the region.

The use or addition of synthetic data in reflectance clustering databases is fairly scarce in the literature; these data are used more often for algorithm derivation, testing and tuning (e.g. Maritorena et al., 2002; IOCCG, 2006; Kuchinke et al., 2009; Gilerson et al., 2010). Cococcioni et al. (2004) created synthetic data for MERIS channels based on

generalized radial basis function neural networks (Cipollini et al., 2001); they then used FCM clustering and fuzzy rules to create models for estimating the concentrations of optically significant constituents in Case 2 waters. Ressom et al. (2006) also incorporated synthetic data from IOCCG (2006) into their clustering database together with NOMAD data; they then applied the FCM clustering approach to train a radial basis function network, which in turn was used to create an ensemble of artificial neural networks.

Synthetic data can be used to supplement clustering databases which contain limited data; tailoring synthetic data according to the spectral characteristics of the smaller clusters (in order to increase cluster size) aids in stabilizing the covariance matrices compared to those of naturally occurring waters, resultantly improving classification success. This could be seen in S1C1 and S2C2 (in figures 3.1 and 3.2 respectively), both representing fairly similar mean R_{rs} and low [Chl a] waters of 61 and 149 spectra respectively, where the average classification success between these two classes improved by 292 % for the South African east coast - a region known to have significant occurrence of near-oligotrophic waters (Smith et al., 2013).

Synthetic data are also useful for creating missing clusters, as seen in steps 2 and 3 with the addition of very low biomass and elevated reflectance clusters. It is vital that the user be aware of the effects that these data have on the changes in the shapes and causality of the dominant modes of variance of the training data set, since these can ultimately affect the clustering algorithm results. A possible concern may be that the addition of synthetic data may distort clusters into groupings of data that do not represent any naturally occurring OWTs. This could potentially be avoided by utilizing representative satellite images with known in-water conditions (e.g. a phytoplankton bloom or flood event) to test whether the clusters are actually classifying these areas according to expectations. If one or more clusters consist of predominantly synthetic data and are not classifying well, those data could be discarded from the database, which may avoid distortion of the mean spectral cluster shapes away from the naturally occurring reflectance spectral shapes. Although the details were not documented in the results, this step was effectively taken during initial testing phases of adding Case 2 synthetic data; results from the calcareous sand and yellow clay mineral type models were excluded from this study since they formed separate clusters and produced spectral shapes that were absent from the regional satellite test images.

Splitting the clustering database based on distinct optical causality

As discussed in section 3.3.1 one of the primary aims for OWT classification in this study was the differentiation of various levels of phytoplankton biomass and subsequent biomass-optimized algorithm application. However, the FCM algorithm tended to create cluster divisions based on magnitude-driven intra-spectral variability, whilst it did not differentiate biomass related OWTs at overall lower R_{rs} spectra caused by absorbing phytoplankton dominated waters. Although it is possible to increase the number of clusters until the low reflectance classes are fully resolved, this would be done at the expense of classification success of the elevated reflectance clusters, since these clusters would decrease in size and become increasingly specialized, which would ultimately jeopardize covariance matrix compatibility compared to those of natural waters.

The current study implemented a database separation based on the amplitude of R_{rs} as indicated by the 560 and 490 nm bands. The separation provided results similar to integral normalization by removing the first order variability in reflectance spectral amplitude linked to b_{bs}; this allowed the clustering algorithm to focus instead on changes in spectral shape, which improved resolution of biomass related OWTs within the lower reflectance phytoplankton dominated waters. The proposed database separation also has similarities to the classical Case 1 and Case 2 approach as defined by Gordon and Morel (1983) and Morel (1988), since the resulting moderate reflectance database resemble Case 1 waters in that the optical variability within the database are primarily determined by phytoplankton; all the classes resulting from this database could be spectrally related to [Chl *a*] via band ratio algorithms, whether in the blue-green or red-NIR spectral regions. The database separation also separated these moderate reflectance data from spectra that resembled classes 1 and 4 from Mélin and Vantrepotte (2015), which were classes that could not be resolved by standard Case 1 water models. The division thus effectively separated the phytoplankton-dominated waters from moderate to highly scattering waters that may require model-based approaches such as inversion schemes accounting for multiple IOPs to resolve concentrations of water constituents.

Fuzzy membership classification is highly amenable to the combined use of clusters from different databases as demonstrated in Moore et al. (2012), where an additional coccolithophore class (created from extracted satellite data) was used in conjunction with previously created OWTs from Moore et al. (2009) (created from a global *in situ* dataset). This is similar to the current body of work, where the classes from the moderate and elevated databases were combined to form the South African OWTs. R_{rs} spectra falling

on the upper or lower edge of the moderate or elevated databases respectively could potentially have partial membership to class 7 (the lowest of the elevated reflectance classes) in addition to some of the moderate reflectance classes. The database separation modification is a useful and functional method for the classification of standard Case 1 versus moderate to highly scattering waters, as long as the optical properties of the resulting clusters are well understood and the represented OWTs tend to exist in natural waters.

The effects of adding satellite data

A number of studies have utilized extracted satellite data to generate or extend OWT classification schemes. González Vilas et al. (2011) extracted satellite water-leaving reflectance data from 15 full resolution MERIS images which had coincident *in situ* match-up data; 150 spectra covering 6 years were available for FCM clustering and the resulting clusters were used as input for multilayer perceptron artificial neural networks for [Chl *a*] retrieval. Mélin and Vantrepotte (2015) utilized SeaWiFS MLAC (Merged Local Area Coverage) data for their global extraction of satellite data from the coastal domain; the extractions were dispersed spatially and over the calender year to ensure optimal coverage for all regions and seasons and resulted in a final clustering training set of 51 million spectra. Moore et al. (2012) extended their existing OWT classification scheme by adding a coccolithophore bloom class consisting of extracted SeaWiFS data; although approximately 600 000 spectra were extracted from 14 images from six different regions spanning six years, only 2500 spectra were subsampled in order to be comparable to the original clustering dataset. A similar step was taken in the current study, where only 900 satellite spectra were added to the clustering database (consisting of 1028 spectra) to avoid outbalancing the signal of the *in situ* data during the clustering process.

The data added to a clustering database can affect the dominant modes of variance of the training set and thus influence the configuration of the resulting clusters; careful consideration is therefore required when specifying the criteria for satellite data extraction and subsampling. In the current study the extractions were performed on 24 images spanning 8 years, comparable to the extraction coverage of previous studies (González Vilas et al., 2011; Moore et al., 2012). Extractions were performed on images classified with clusters from dataset 4, as these clusters represented some of the most appropriate OWT shapes and [Chl *a*] ranges for the desired classification applications. The primary criteria was that only poorly classified pixels (membership <0.1) with valid spectra were extracted in order to find missing spectral shapes, and that these extracted data be

separated into the moderate and elevated groups as discussed in the previous section. Another important criterion, based on the intended classification applications for satellite data, was the continued use of the division of the moderate database into groups of increasing [Chl *a*] ranges. This ensured that possible missing spectral shapes were included whilst preserving the foundation of the existing cluster shapes; it also facilitated the creation of a new cluster (S5C6M shown in figure 3.5 in section 3.2.1) aimed at the classification of very high biomass phytoplankton blooms in the southern Benguela.

A crucial premise in the successful application of OWT classification is the assumption that the remotely sensed satellite reflectance closely resembles the actual water-leaving reflectance. However, this is often not the case in turbid or productive coastal waters where atmospheric correction failure often leads to erroneous or negative water-leaving reflectance; an example of this is given in figure 3.20 of a MODIS spectrum extracted over a high biomass bloom. The reliability of the satellite measured reflectance is greatest if there is appropriate consideration of the various sources of error and sensitivity of the satellite retrieved reflectance, and this information is assessed together with the knowledge of the study area (Blondeau-Patissier et al., 2014). The bright pixel atmospheric correction (BPAC, Moore and Lavender, 2011) is very often, and in the example given in figure 3.20, ubiquitously applied over the shelf waters of the west coast. This atmospheric correction has been shown to be largely appropriate for the southern Benguela by generally maintaining spectral features associated with phytoplankton blooms (Evers-King, 2014); however, BPAC can produce a spectral offset compared to *in situ* measurements, as shown in the example in figure 3.20, that could result in poor classification of satellite imagery. It is difficult to distinguish whether the OWTs are not well represented from a bio-optical perspective, or if poor classification is just an artefact of atmospheric correction failure. Although the assessment of a variety of different atmospheric correction techniques is not within the scope of this thesis, it is vital to assess the potential impact of the standard MERIS atmospheric correction on the OWT classification ability; this will be discussed further in chapter 4.

The primary benefit of the inclusion of satellite data is the consequent optimization of the clustering database for the radiometric characteristics and performance of that particular sensor. By using the R_{rs} data from a particular sensor, some of the spectral ambiguities caused by the atmospheric correction are effectively incorporated into the classification. Take for example figure 3.20, where the shape of the MERIS spectrum is very similar to the *in situ* spectrum with a spectral offset of approximately 0.001 sr^{-1}.

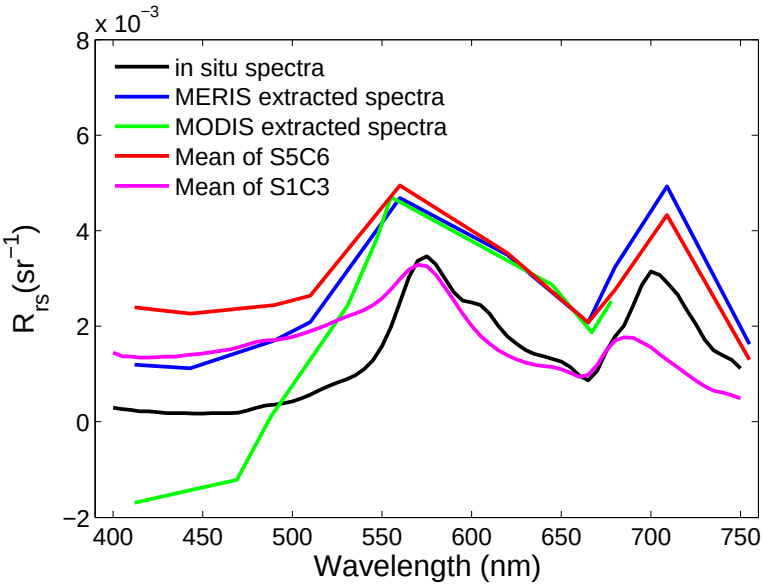


Figure 3.20: The R_{rs} spectra of an *in situ* (black) and MERIS (blue) and MODIS (green) satellite extracts for 30 March 2005, compared to the spectral mean of class 6 from step 5 (red) and class 3 from step 1 (magenta). The *in situ* data show an example of a *Prorocentrum triestinum* (small dinoflagellate) bloom in the southern Benguela, from Bernard et al. (2014).

If classifying with a cluster created with only *in situ* data (class 3 from step 1 as seen in figure 3.1), the satellite spectrum has very low membership to this class ($f_i = 0.00002$), whilst even the *in situ* spectrum has a membership of only $f_i = 0.0027$. However, when classifying with a class that contains predominantly satellite data (class 6 from step 5 from figure 3.5), the satellite spectrum has a membership of $f_i = 0.587$ to that class. Since both class 3 and class 6 have a similar size ($N = 107$ and $N = 136$ respectively) it is most likely the shape of the mean and covariance matrices of the clusters that affect the classification success. Although the addition of synthetic data with very high [Chl *a*] (e.g. Robertson Lain et al. (2014)) could potentially aid in the creation of a hyperspectral class with a spectral shape more similar to the satellite spectum (causing higher reflectance peaks in the green and red), this still would not account for the offset in magnitude of the satellite data. The addition of satellite data with similar spectral shapes and offsets, and potentially known optical properties, are a pragmatic solution to the classification problem.

In this study there is also a regional optimization aspect that should be considered. Since EAP data were only created for the southern Benguela, the MERIS satellite data were only extracted from this region. This may have some inherent effects on the classification success in other regions. Apart from possible differences in regional in-water bio-optical properties, there may also be changes in the atmospheric properties between

the southern Benguela and the other coastal regions of South Africa, including the dust, aerosol and water vapour content of the atmosphere. Evidence of decrease in regional classification success could be seen in the table 3.1 where the average classification success decreased in each of the training datasets from west to east along the coast of South Africa. The classification success and spatial and temporal persistence of the OWTs will be discussed in more detail in chapter 4.

3.3.3 Transferability of the methods

Transferability to other sensors

Although the addition of multispectral satellite R_{rs} in this study limits the overall application of the final clusters to only one sensor (MERIS), the general approach can still be applied using data from other ocean colour satellites. The use of hyperspectral *in situ* and synthetic data allows the clusters from datasets 1 to 4 to be resampled to the spectral resolution of another sensor, e.g. MODIS. By applying the appropriate criteria the extraction of satellite data could be done after classification with clusters resulting from either dataset 3 or 4, depending on the desired application and whether a database separation was required. Unfortunately this would negate a direct comparison of the present study's classification results to those from different sensors, as done in Moore et al. (2014) for both MERIS and MODIS, since the resulting clusters would most likely represent different OWTs. A qualitative comparison of OWT presence and persistence between sensors may be possible, but would require comprehensive knowledge of the bio-optical characteristics of the OWTs and the effects of atmospheric correction on image classification.

For use in the productive waters of the southern Benguela, a sensor with the appropriate bands in the red-NIR regions would be desirable. For future applications the classification could potentially be applied directly to data from the Ocean and Land Colour Instrument (OLCI) aboard Sentinel-3; OLCI is based on the heritage of MERIS with a similar spectral band configuration and algorithm types.

Transferability to other regions

Since the original *in situ* data and satellite data were obtained from the southern Benguela, the clusters and resulting OWTs are considered to be optimized for the atmospheric and in-water conditions inherent to that region. Evidence of this can be seen in table 3.1 where the average classification success decreased towards the east around the coast of South Africa. The representativeness of these OWTs in South African waters will be assessed further in chapter 4. Theoretically the classification with the OWTs described in this chapter may also perform well globally in regions with similar bio-optical conditions and atmospheric correction performance to the southern Benguela, such as other western boundary upwelling zones.

3.3.4 Applicability of inversion and Chl *a* algorithms

The use of the EAP algorithm to derive IOPs

The sole use of [Chl *a*] is often not enough to define or characterize OWTs in coastal waters since the same [Chl *a*] can represent R_{rs} of various shapes and amplitudes; given the ambiguity in R_{rs}, the composition of IOPs could provide additional descriptive qualities. In the absence of *in situ* measured IOPs the Equivalent Algal Populations (EAP) inversion algorithm, as outlined in section 2.3.1, was applied to *in situ* and satellite R_{rs} data to derive IOPs for determining the causality of the dominant modes of variance of the clustering databases and final OWTs; the EAP was chosen since it was designed for, and has been validated in, the phytoplankton-dominated waters of the southern Benguela (Evers-King et al., 2014), and provides considerable IOP information.

The EAP [Chl *a*] product appears to be less prone to overestimation than the MPH [Chl *a*], the other regional product available for the southern Benguela (Evers-King, 2014), particularly at high biomass conditions; it also successfully retrieves [Chl *a*] over a larger range than other available ocean colour products (see Matthews et al., 2012 for discussion of the performance of other products in the southern Benguela). The greatest capacity for the EAP inversion technique to find the correct solution in terms of [Chl *a*] is under conditions of low a_{dg} and b_{bs} when phytoplankton absorption and scattering become dominant components in the IOP budget (Evers-King, 2014). This is generally true for much of the southern Benguela, since a_{dg} from riverine influxes are limited by the dry climate along the west coast of South Africa, whilst sediments are generally constrained to the shallow inshore and bay regions where resuspension may occur. The satellite-derived [Chl *a*] and modelled IOPs produced by the EAP should be reliable enough for causality analysis for most of the moderate reflectance classes.

The EAP employs a relatively simple constant parameterization for the absorption of CDOM and detritus, which can lead to the lower convergence and the higher errors in the blue (between 400 and 430 nm) when applying the algorithm to *in situ* and modelled data (Evers-King, 2014); this may be due to the low radiance values found in this spectral region in highly absorbing waters, where the ambiguity in predicting the total absorption in ocean colour inversion can be high (Defoin-Platel and Chami, 2007). The reliability of the EAP [Chl *a*] product has also not been established by validation with *in situ* data under conditions where phytoplankton are not the dominant contributor to the IOP budget. Under conditions of higher biomass and increased b_{bs} the assumptions associated

with the reflectance approximation (as applied to satellite data) begin to break down (Evers-King, 2014), making the resulting IOPs and other products less reliable; caution is thus recommended when using the IOPs and [Chl *a*] from the EAP algorithm for the high reflectance classes. For this reason the standard MERIS Case 2 product Algal2 data were extracted instead of the EAP chlorophyll product for the elevated reflectance classes. However, even if the causality of variance of the high reflectance OWTs is unclear these classes can still be useful for classification purposes solely based on their magnitude and spectral characteristics; these classes indicate moderate to highly scattering waters that potentially contribute to algorithm error (i.e. areas where Case 1 algorithms would fail). Since the IOPs and [Chl *a*] from the EAP algorithm were only used in a descriptive manner, and not to parameterize or develop new bio-optical algorithms, the potential errors introduced via algorithm and data constraints are acceptable.

Class-specific Chl *a* algorithms

Although the use of specific spectral shapes or features can be sufficient to discriminate between various OWTs, this method often yields only qualitative or phenological results (Blondeau-Patissier et al., 2014); it is the bio-optical algorithms that are applied to these classified reflectance spectra which typically provide the quantitative values for water constituents. Due to the paucity of regional *in situ* data for the development and parameterization of new class-specific algorithms (as done in Moore et al., 2001; Le et al., 2011), the focus in the current study was rather to analyse the performance of an assortment of existing band-ratio algorithms, and to select the most appropriate algorithm per OWT (as done in Moore et al., 2014).

It is important to consider the various sources of uncertainty affecting the derivation and satellite application of the respective chlorophyll algorithms applied in this study. Dierssen (2010) noted that much of the variability found in the empirical data used to create Chl *a* algorithms may stem from errors in radiometric and [Chl *a*] measurements. Algorithms that are derived from *in situ* R_{rs} would incorporate the uncertainties and errors inherent to the *in situ* radiometric measurements (e.g. calibration errors and self shading) and processing (e.g. subsurface extrapolation methods); these errors can be affected by water type and environmental conditions (Zibordi and Voss, 2010) and tend to have varying magnitudes over the visible spectra, generally with the greatest uncertainties in the red (Zibordi et al., 2004; Zibordi and Voss, 2010). Even the uncertainty associated with the derivation of fluorometric chlorophyll, often used in the process of relating radiometric data to *in situ* [Chl *a*], may be in the order of 50 % (Trees et al., 1985). These errors and

uncertainties need to be taken into account when choosing class-appropriate chlorophyll algorithms per OWT.

The statistical tests of the class-specific algorithm application (as shown in tables 3.3 to 3.6 in section 3.2.6) only included a portion of the data that make up the mean spectral shape and covariance matrix of the individual OWTs used for satellite image classification; resultantly, these error and uncertainty results were primarily used as an indicative tool for selecting the most appropriate algorithm per OWT for the weighting and blending procedure when applied to satellite R_{rs}. The application of this classification framework to satellite matchups (as done in section 3.2.7) provides better estimates of the errors and uncertainties associated with these algorithms when applied to MERIS R_{rs}, which can in turn be utilized to provide (preliminary) blended uncertainty products to accompany blended [Chl *a*] products (as shown in figure 3.19 in section 3.2.8).

1. Low to medium biomass conditions

The blue-green band ratio algorithms showed the best performance within the low to medium biomass classes 1 to 4, particularly at [Chl *a*] below 10 mg m^{-3}; these types of algorithms have previously been applied (although not specifically validated) during studies of the southern Benguela (Machu et al., 1999; Demarcq et al., 2003; Weeks et al., 2006; Silió-Calzada et al., 2008). Smith et al. (2013) have validated the performance of Algal1 in the Case 1 waters of the KwaZulu-Natal bight on the east coast of South Africa, where it was shown to have an absolute percentage difference of 54 % compared to *in situ* [Chl *a*] ranging from 0.35 to 1.05 mg m^{-3}. The OC4E algorithm has been shown to have a RMSE of 0.323 and absolute percentage difference of 38.76 % when tested on a SeaBASS dataset (N = 644) covering an *in situ* [Chl *a*] range of 0.001 to 40.2 mg m^{-3} (Feldman and McClain, 2015); this is comparable to the [Chl *a*] ranges covered by the low to medium biomass classes 1 to 4 in the current study. The relatively high MARE values (70.7 - 121 % for OC4E, as shown in table 3.6) for these classes are mostly attributed to the optical variability introduced by the synthetic data, since the MARE of the *in situ* data were generally lower (<55.4 %). However, these ranges are still comparable to results from Moore et al. (2014) where the OC4 algorithm showed MARE of 41.2 - 109.8 %. Two of the Moore et al. (2014) OWTs (OWTs 3 and 2) also had mean spectral shapes and [Chl *a*] that resembled OWTs 4 and 5 from the current study; an OC4-type algorithm showed the best performance for these OWTs in both studies.

The variability in the performance of the algorithms depends on the similarity of the spectral characteristics (and hence the IOPs) of each class compared to the training data of each of the algorithms. The range of application for Algal1 is 0.01 to 30 mg m^{-3} (Morel and Antoine, 2011), which is slightly lower than the range of the NOMAD dataset that was used to train OC4E (0.012 to 72.1 mg m^{-3}, from Werdell and Bailey, 2005); the difference in training ranges may explain why Algal1 showed lower RMSE and MARE for the low biomass classes 1 and 2, whilst OC4E and OC3E performed better in the low to medium biomass conditions of classes 3 to 5. The differences in statistical performance between the blue-green algorithms for the low to medium biomass classes demonstrates the algorithms' sensitivity to optical variability even at relatively small [Chl *a*] ranges.

2. High biomass and bloom conditions

Many of the bio-optical algorithms designed for more eutrophic conditions found in coastal environments and lakes have opted for spectral algorithms using the red-NIR bands (Hu et al., 2010; Matthews et al., 2010; Gitelson et al., 2011b; Yacobi et al., 2011; Moore et al., 2014); in the southern Benguela the MPH (Matthews et al., 2012; Smith and Pitcher, 2015) and EAP (Evers-King et al., 2014) algorithms, as well as a version of the two band 709/665 algorithm (Bernard et al., 2005), have been applied to MERIS data and have generally performed better than the standard blue-green band ratio algorithms.

The results from classes 5 and 6 were similar to Moore et al. (2014) where the red-NIR band ratio algorithms showed the best performance for [Chl *a*] above 10 mg m^{-3}. The general format of the 2 and 3 band red-NIR models used in the current study have previously been successfully applied to *in situ* (Yacobi et al., 2011), synthetic (Gilerson et al., 2010) as well as satellite data (Moses et al., 2009b; Gitelson et al., 2011a); however, these models can be susceptible to natural variability of the Chl *a* specific absorption coefficient of phytoplankton (a^*_ϕ) as well as the quantum yield of Chl *a* fluorescence (Gurlin et al., 2011) and may thus require regional parameterization. Analytically developed advanced versions of these algorithms, specifically designed to be used without regional tuning (Gilerson et al., 2010), were applied to class 5 and 6; the data ranges of the two OWTs corresponded to the relatively high [Chl *a*] used to create these algorithms. However, studies have found that the generalized relationship between a^*_ϕ and [Chl *a*] changed rapidly at [Chl *a*] below approximately 10 mg m^{-3} (Gilerson et al., 2010; Shi et al., 2013). These algorithm limitations manifested in the form of negative returns and increased scatter of the G2B and G3B modelled data at *in situ* [Chl *a*] <10 mg m^{-3} (figure

3.16). Previous studies have recommended a minimum threshold of approximately 10 mg m^{-3} [Chl *a*] for these types of 2 or 3 band red-NIR algorithms (Dall'Olmo and Gitelson, 2006; Moses et al., 2009a,b). Resultantly, a switching algorithm was implemented which would use OC4E as the default product, and only switch to the G2B or G3B algorithms above a threshold of 10 and 15 mg m^{-3} respectively. This technique improved the RMSE and MARE for the blended product in both the *in situ* and the satellite matchup datasets, and helped to facilitate the optimal utilization of data and algorithms over various water types with highly dynamic biomass ranges.

3. Elevated reflectance classes

None of the blue-green band ratio algorithms performed adequately in the elevated reflectance classes; although the red-NIR algorithms statistically showed better performance, there was still a great deal of data lost to negative returns, possibly due to the low phytoplankton biomass signal, making them generally unsuitable for application to these classes. For satellite application the Algal2 product was applied to the high reflectance OWTs due to the convenience of it being the standard Case 2 product distributed with MERIS data. Since it was not possible to test the algorithm performance on the *in situ* and synthetic data it is difficult to comment on whether this would actually be the best algorithm for the high reflectance OWTs. Nevertheless, Algal2 proved to be more robust than the blue-green algorithms in the southern Benguela validation dataset and performed adequately for the MERMAID database where 55 of the 241 spectra were classified as elevated reflectance OWTs 7 to 11.

A primary focus of this body of work was on assessing the performance of known and well-documented [Chl *a*] algorithms within the OWT classification framework. Although many inversion algorithms exist that could also potentially be applied to the elevated reflectance classes, these algorithms would most likely require regional parameterization, which falls outside the scope of this thesis. This could even be done for the Algal2 product, which is derived from the absorption coefficient for all pigments (a direct output of the neural network) by means of a simple equation; within the ODESA processing framework, which provides the source code of the MERIS Level 2 ground segment (MEGS) and future sensors on board Sentinel 3, it is possible to adjust the parameters within the Algal2 equation according to regional values, if available.

Potential uses for these OWTs include the classification of regions where the standard Case 1 algorithms should be avoided. These regions are also indicative of areas that could benefit from increased *in situ* data collection, specifically for waters with increased b_{bs} and/or a_{gd}; these *in situ* data would aid the development and parameterization of regional or class-specific algorithms for the South African elevated reflectance classes. Newer and more appropriate algorithms can easily replace Algal2 within the algorithm blending framework as they become available.

4. Combined results for the blended Chl a product

As in Moore et al. (2014) the blended [Chl *a*] product gave improved RMSE and MARE results compared to the individual algorithms for the combined *in situ* and synthetic datasets. The recommended switching setup for classes 5 and 6 further improved results for the satellite matchup datasets. In all cases the blended product (utilizing the switch) showed lower RMSE compared to using any of the single algorithms.

The desired satellite product accuracy goals for open ocean waters has been broadly defined as approximately 35 % (McClain, 2009; IOCCG, 2012). When applied to global satellite matchup data, the algorithm blending procedure implemented in this thesis produces a MARE of approximately 45 % (similar to 47.9 % from Moore et al. (2014)) over a [Chl *a*] range spanning almost four orders of magnitude. The blended product produces better results for the global satellite matchup dataset than achieved by any of the individual global algorithms tested in this study; when applied to only *in situ* data, the blend also shows better MARE results for the southern Benguela (38 %, from table 3.6) compared to the locally-tuned EAP algorithm, which had an absolute percentage error of 73 % (Evers-King, 2014) relative to *in situ* [Chl *a*]. The algorithm blending procedure applied in the current study was tested on a variety of radiometric sources, and the results support the concept that an approach which blends and switches between algorithms fine-tuned to specific concentration ranges provides better retrievals than applying one algorithm to resolve a broad and dynamic range of environmental conditions (Moore et al., 2014; Matsushita et al., 2015).

3.3.5 Conclusions and recommendations

The objective of this chapter was to provided a provisional optical characterization of the OWTs prevalent in the coastal waters of South Africa in order to facilitate:

- the direct identification of different bio-optical regions within a satellite image independent of a derived chlorophyll product.
- the classification of satellite data for the identification of specific OWTs as standalone products (e.g. high biomass phytoplankton blooms, or turbid coastal waters).
- The assessment and selection of appropriate class-specific chlorophyll algorithms for application through an algorithm blending framework, for the purpose of creating an optimized Chl *a* product.

User intervention and regional knowledge are required to evaluate the clusters' ecological significance and classification success, and to ensure that the final classes meet the clearly defined objectives.

The following innovative methods were introduced in this study:

- To the best knowledge of the author, this is the first study that combined the use of *in situ*, synthetic and extracted satellite R_{rs} for creating optical water types. It has been shown in the current study that the addition of different data sources to the input database used in the clustering process had a significant effect on the dominant modes of variance, which subsequently impacts the resulting clusters' shape and the classification success (whether statistically, or subjectively based on user objectives).
- A database separation method is proposed in this study; the method separates the phytoplankton dominated waters from moderate to highly scattering waters. The resulting moderate reflectance database offers similar advantages to an integral normalized database, which allows the clustering algorithm to focus on the spectral shape, whilst also allowing for the optimal distinction of different biomass levels in phytoplankton dominated waters. The spectral amplitude has a more dominant effect in determining the formation of the elevated reflectance classes.

The conclusions that can be drawn from this study:

- The addition of synthetic data to the clustering database is useful not only for the creation of missing clusters, but also by increasing the knowledge base of the optical causality that leads to similar spectral shapes. Although the increased optical variability can increase the RMSE and MARE, particularly in the phytoplankton dominated classes when using blue-green and red-NIR band ratio algorithms, it remains useful to quantify the uncertainties that can be associated with R_{rs} ambiguities within OWTs.
- The relatively low classification success of the datasets with solely *in situ* and synthetic data, compared to those optimized by satellite data, indicate that there are discrepancies in the R_{rs} between these data sources; this highlights the importance of understanding quantitative effects of the atmospheric correction on classification ability on a per-sensor basis.
- Due to regional *in situ* and satellite data constraints, the current set of OWTs are optimized for the coastal waters of the southern Benguela. Classification success appears to lessen from the west towards the east coast of South Africa; the frequency and spatial extent of low total membership to the current set of OWTs will be addressed further in chapter 4.
- The addition of regionally extracted satellite data to the clustering database may potentially limit the global applicability of the these OWTs; their regional extent and applicability need to be tested in different atmospheric and bio-optical conditions before application of the blended Chlorophyll product.

Some of the following recommendations can be made:

- In this study the focus was only on testing the performance of a selection of band ratio algorithms, consisting of three blue-green and two red-NIR algorithms, and one neural network output. Future studies could include more specialized semi-analytical or other model-based approaches tailored to the optical properties of the specific OWTs.
- The successful classification of satellite imagery relies heavily on the ability of the atmospheric correction to return accurate reflectance data. Further investigation of appropriate atmospheric correction techniques is required for the southern Benguela region in order to provide more quantitative effects of the atmospheric correction choice on the classification performance and associated algorithm application.

- The collection of more *in situ* sensor-independent hyperspectral radiometric data, with coincident biogeochemical parameters and IOPs, would facilitate a more comprehensive and authentic regional OWT characterisation. Data collection should focus on the inclusion of low biomass conditions, as well as optically complex coastal regions affected by suspended sediments and coloured dissolved organic matter.

The present set of OWTs are not finite and there is potential for intermittent revision of the clustering database as new *in situ* data become available; this process would simply require re-clustering (with or without the inclusion of extracted satellite data) to define a new set of cluster means and covariance matrices. The adopted classification framework, as pioneered by Moore et al. (2001, 2014), is also amenable to alteration by substitution of more favourable alternative class-specific algorithms that may not yet have been tested in the study region.

The OWTs developed and defined in this chapter offer substantial utility for time series analysis of different ecological provinces in the coastal waters of the southern Benguela and the rest of South Africa. The application of these OWTs will be demonstrated in the next chapter by utilizing them to classify the entire MERIS archive for South African coastal waters to show the OWT phenology and persistence, in addition to creating blended [Chl *a*] products.

Chapter 4

Optical water type application: classification of MERIS imagery

4.1 Introduction

Ocean colour remote sensing is a powerful tool that provides data at spatial and temporal scales unattainable by standard *in situ* methods. The utility of satellite ocean colour data for studying geochemical and biological processes in oceanic, coastal and inland waters is well established; there are decades of archived mid-resolution satellite ocean colour data available for wide ranging applications such as analysing trends in primary production (e.g. Petrenko et al., 2013; Kahru et al., 2015), phytoplankton dynamics (e.g. Blondeau-Patissier et al., 2014 and references therein), and water quality (e.g. Matthews, 2014). Each of these applications has specific spatio-temporal data requirements, whilst the optical nature of relevant target bodies necessitates application-optimized algorithms that, when applied to satellite data, can resolve the biogeochemical constituents within suitable uncertainty limits.

The coastal zone is particularly challenging for ocean colour application since the in-water conditions can vary massively in optical complexity and trophic state over small spatial scales, necessitating algorithms that can deal with highly dynamic ranges of biogeochemical constituents. It has been demonstrated that an approach that can pre-classify different optical water types and subsequently apply and blend class-specific algorithms, tailored for a specific range of IOPs or biogeochemical constituents, is more powerful and accurate than the application of one algorithm over a dynamic range of phytoplankton biomass and in-water optical conditions (Moore et al., 2014). Chapters 2 and 3 have detailed a classification framework for the identification and characterization of a variety

of OWTs that may be found over the continental shelf waters of South Africa, and the subsequent class-specific resolution of [Chl a]. This chapter will explore the utility of this classification system through direct application to satellite water-leaving reflectance data.

4.1.1 Considerations for the utility of remotely sensed data

The utility of ocean colour remote sensing is application driven and it is important that the chosen approach (e.g. the sensor's spectral band configuration, radiometric performance, revisit time, etc.) complements the specific requirements of the study. Different OWTs inherently have distinct ranges of reflectance signals and resulting spectral radiometric requirements; whilst blue-green algorithms may be enough to resolve constituent concentrations in low biomass phytoplankton-dominated waters, eutrophic and Case 2 waters may require large signal variability and error considerations across the visible spectrum.

In terms of spectral configuration: the studies focussing on regional and global optical characterization (e.g. Vantrepotte et al., 2012; Mélin and Vantrepotte, 2015) have been able to successfully utilize the spectral region between 412 and 670 nm provided by SeaWiFS and MODIS, while studies within turbid productive and phytoplankton dominated waters have benefited from the 709 nm band offered by MERIS (e.g. González Vilas et al., 2011; Moore et al., 2014; Shen et al., 2015). As in Moore et al. (2014) the OWT classification method employed in this thesis utilizes MERIS wavebands 2 - 10 (between 443 and 753 nm); thus it is necessary that the satellite reflectance data that falls within this spectral region be of sufficient quality for successful image classification, as well as successful application of chlorophyll algorithms.

In terms of data quality: atmospherically corrected level 2 reflectance data of sufficiently low uncertainty across the visible spectrum is crucial to ensure consistent performance of the OWT reflectance classification algorithm. Errors and uncertainty may be introduced at various points in the processing chain, from the instrument/sensor itself to the methodology used for the corrections. Sources for uncertainties in satellite retrieved normalized water-leaving radiance measurements may include pre-launch calibration, bi-directional correction, atmospheric correction, and in-orbit changes in the sensor's performance, whilst additional uncertainty may also be introduced by geo-location errors and land adjacency (IOCCG, 2006). The water leaving signal that is measured by the sensor at the top of atmosphere (TOA) is at most approximately 10% (Antoine et al., 2008), whilst the majority of the TOA signal is from the atmosphere and a combination of physical contributions from other sources, i.e. those radiances from sun glint, whitecaps,

air molecules, aerosols, and the various interactions between them (IOCCG, 2010). Odermatt et al. (2012) noted the importance of appropriate atmospheric correction capable of preserving the NIR reflectance peak in optically complex water. For MERIS the outperformance of the bright pixel atmospheric correction (BPAC, Aiken and Moore, 2000), used in the standard processing chain (although the 3rd reprocessing uses Moore and Lavender, 2011), over the Case 2 regional (C2R, Doerffer and Schiller, 2008) atmospheric correction in productive waters has been demonstrated by Moses et al. (2009a). Other studies have noted the failure of the C2R atmospheric correction to retrieve the NIR reflectance peak (Guanter et al., 2010; Odermatt et al., 2010 cited in Odermatt et al., 2012). Strong water absorption in the NIR results in low magnitudes of reflectance in this spectral region, which can lead to increased noise in the measured signal (Gitelson et al., 2008). Requirements for satellite water-leaving reflectance products were put as <5% in oligotrophic open ocean waters, although this goal is often not attained for all wavelengths (Antoine et al., 2008; Mélin et al., 2016), and has been shown to vary between optical water types and sensors (Moore et al., 2015). Validation of satellite radiometric and biogeochemical products is necessary to ensure that these products are of sufficient quality to meet the user needs that they were designated for. Various validation (e.g. Werdell and Bailey, 2005; Antoine et al., 2008) and vicarious calibration (e.g. Clark et al., 2003) activities are in place for satellite missions to assess and characterize the sensor-specific errors and uncertainties (IOCCG, 2006). Quality flags are activated at various steps of the satellite data processing chain if data are not of a sufficient standard.

In terms of temporal and spatial resolution: coastal regions are generally more dynamic and optically variable than open ocean environments (Davis et al., 2007). Satellite data needs to be able to provide sufficient spatial and temporal coverage to resolve the feature(s) of interest (e.g. phytoplankton blooms, river plumes) and their inherent variability; for the coastal ocean, Moses et al. (2016) has suggested sensor spatial resolutions of 200 m or less in order to avoid subsuming small-scale features within satellite pixels. Thus the scales of spatial heterogeneity that may occur within a study region needs to be taken into account; this is particularly important in regions with a high degree of horizontally patchy surface algal or cyanobacteria blooms where certain satellite sensors may not be able to resolve the high temporal and/or spatial variability of the blooms (e.g Kutser et al., 2006). Uncertainties with regard to patchiness associated with comparisons between *in situ* samples and satellite data can be in the order of 11% (Matthews et al., 2012) in the southern Benguela. Uncertainties also arise from vertical inhomogeneity of the water column where comparisons are made between *in situ* samples from discrete

depths, and satellite measurements representing a weighted average of the near-surface layer of the ocean (Gordon and Clark, 1980); this is of particular importance during periods when phytoplankton biomass maxima occur subsurface (e.g. Pitcher and Cockcroft, 1998; Kutser, 2004). Regional knowledge is needed about when and where these types of conditions may occur, and to recognize the limitations of satellite data under these circumstances.

4.1.2 The physical and biological characteristics of South African coastal waters

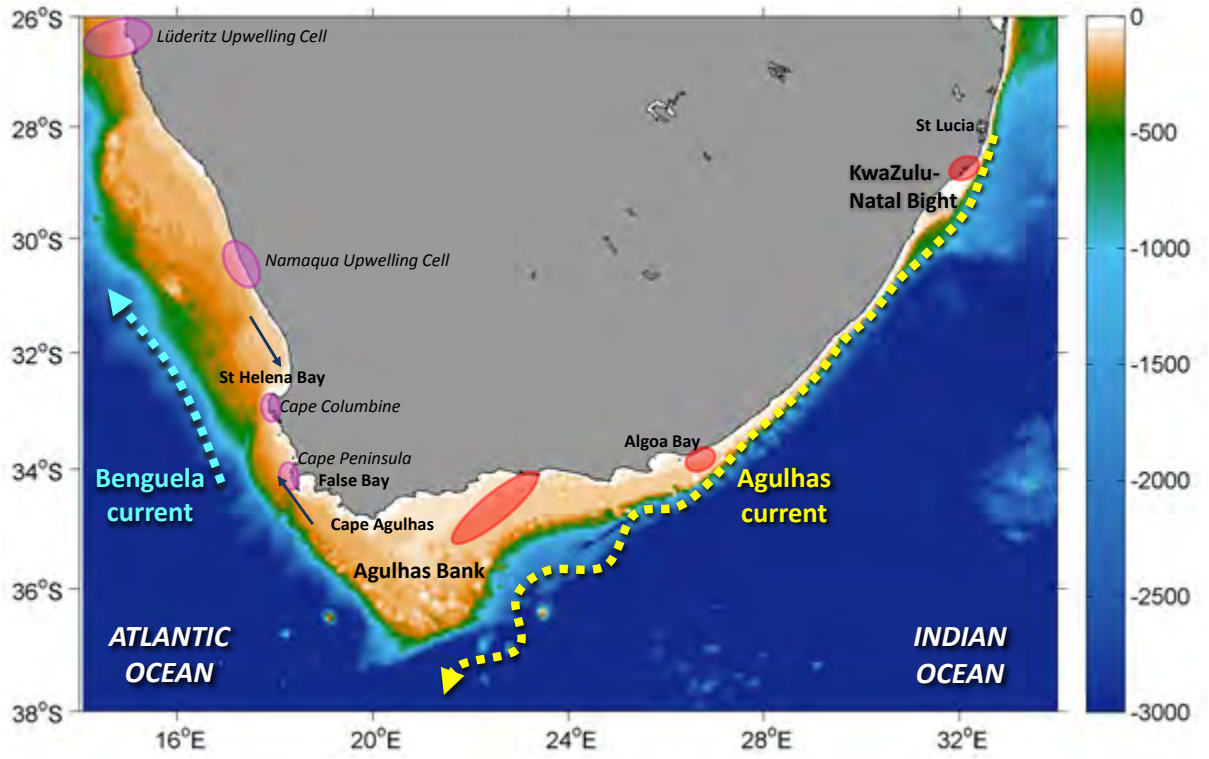


Figure 4.1: Schematic map detailing the bathymetry and key oceanographic features of South African coastal regions. The Benguela upwelling cells are shown in pink, whilst the zones of upwelling and/or periodic nutrient enrichment on the south and east coasts are shown in red.

As its name suggests South Africa is situated at the southern tip of the African continent. The position of the region is unique in that the coastal waters are affected by both the Atlantic and the Indian Ocean; whilst the east coast is flanked by the poleward flowing Agulhas Current, a warm and fast-flowing western boundary current, the west coast is characterized by the colder waters associated with the Benguela upwelling system (figure 4.1). The complex interactions between these two currents and with the underlying bathymetric features together with various atmospheric forcing mechanisms combine to influence the physical characteristics, nutrient availability and the resulting ecological variability along the coast of South Africa.

Much of the western and southern coastal regions of the country forms part of the Benguela Current Large Marine Ecosystem (BCLME), one of the four major eastern boundary upwelling systems, as well as one of the most productive ocean areas (Shannon, 2006), in the world. Only the southern Benguela is applicable to the current body of work;

this region is bordered to the north by the Lüderitz upwelling cell at approximately 26°S (Hutchings et al., 2009), and by the Agulhas Current and retroflection zone (Veitch et al., 2009) to the south. The southern Benguela is primarily driven by equatorward wind stress from the South Atlantic high pressure cell (Shannon, 1985) which produces upwelling-favourable winds during austral summer. Throughout the upwelling season mesoscale atmospheric perturbations such as cold fronts and coastally trapped lows, and even diurnal winds (Lucas et al., 2014), create a pulsed system which is crucial for driving short-term variability in upwelling, nutrient availability and plankton development (Hutchings et al., 2009). Discrete upwelling centres, as indicated by the pink regions in figure 4.1, form within the southern Benguela due to a combination of the coastal topography and shelf width (Hutchings et al., 2009), resulting in increased phytoplankton concentrations downstream of these cells (Verheye et al., 2016 and references therein).

The Agulhas Bank region displays characteristics of both an upwelling and temperate shallow shelf system (Hutchings et al., 2009); whilst wind-induced coastal upwelling may occur periodically off the prominent capes between Algoa Bay and Cape Agulhas during summer (Schumann et al., 1982; Largier et al., 1992), the rest of the Agulhas Bank displays seasonal variability with thermal stratification and deep mixing during summer and winter respectively (Probyn et al., 1994). Other features include a semi-permanent upwelling cell upstream of Algoa Bay (Lutjeharms et al., 2000a) where cool productive water surfaces as the Agulhas Current diverges from the coast (Hutchings et al., 2009), as well as a cool ridge of nutrient-rich water (22°E), formed by a combination of coastal wind-induced upwelling and the Agulhas Current, which is periodically found at the surface stretching from the coast in a south westerly direction (Jackson et al., 2012). These productive waters make the Agulhas Bank both an important spawning ground and nursery area along the South African coast (Hutchings et al., 2002). The Agulhas Current follows the shelf break along the eastern edge of the Agulhas Bank, and influences the oceanography on the shelf through various meanders, filaments and current intrusion onto the outer Bank (Krug et al., 2014b). The Agulhas Current occasionally sheds eddies and filaments northwards into the South Atlantic that may influence the shelf ecosystems (Hutchings et al., 2009).

The physical characteristics and environmental conditions along the east coast are governed by the warm oligotrophic waters of the southern Indian ocean and the fast-flowing waters of the Agulhas Current. The region has a smooth coastal topography, generally narrow continental shelf and steep slope gradient which allows the Agulhas Current to fol-

low a stable trajectory close to the coast (de Ruijter et al., 1999). The only distinct offset along the coastline is the KwaZulu-Natal Bight which has a broader continental shelf and more moderate slope gradient compared to the adjacent coastline. The Agulhas Current tends to follow the 200 m isobath and thus effectively encloses the shelf waters of the bight. A dynamically-driven upwelling cell is located just south of St Lucia along the inner edge of the Agulhas Current where it moves offshore to follow the shelf edge (Lutjeharms et al., 2000b); this upwelling cell brings colder, nutrient-rich South Indian Subtropical Surface Water onto the shelf which is transported southward over the bight at all depths (Lutjeharms et al., 2000b) and along the inner edge of the Agulhas Current (Meyer et al., 2002), and can have a substantial influence on the phytoplankton productivity over the whole bight (Carter and d'Aubrey, 1988; Carter and Schleyer, 1988). Several rivers flow into the KwaZulu-Natal Bight which provide additional nutrients (Carter and Schleyer, 1988; Meyer et al., 2002), as well as inputs of sediments and dissolved organic matter, to the shelf waters. The region is characterized by eddies and re-circulation patterns (Schumann, 1987) which create a mechanism for retaining more phytoplankton biomass relative to the dynamic Agulhas Current region (Carter and Schleyer, 1988), whilst sediments and other particulate matter may also have a longer residence time on the bight than the adjacent more exposed shelf environment (Van Ballegooyen et al., 2007). The retentive nature of the surface waters, as well as various estuarine inputs of freshwater and nutrients, have made the KwaZulu-Natal Bight an important spawning ground (Hutchings et al., 2002).

4.1.3 Why use optical water type classification in South Africa?

From the previous section it is clear that South African coastal waters exhibit a dynamic range of environmental and bio-optical conditions; however the optical characteristics of all these regions have not been well documented, and most field studies have focussed on the productive west coast. Since there is a paucity of *in situ* radiometric and IOP data for the south and east coast regions of the country, the use of satellite classification can help to better characterize these under-sampled areas, whilst also identifying priority regions for future *in situ* sampling efforts.

Blue economy strategies are becoming increasingly important with government initiatives like operation Phakisa (van Wyk, 2015) aiming to unlock the economic potential of South Africa's oceans with a strong marine protection and ocean governance as well as aquaculture focus. The societal benefits of ocean colour for water quality monitoring (IOCCG, 2008) and aquaculture and fisheries management (IOCCG, 2009b) are well established, with accurate quantification of biogeochemical information being a key requirement. Kirkman et al. (2016) acknowledged the importance of collecting and assessing accurate satellite-derived products in the BCLME for utility in ecosystem-based management as it provides the ability to extend and contextualize surface information collected from existing transect lines and *in situ* stations, in addition to supplying quasi-synoptic views of the system.

Ocean colour clearly has an important role to play in South African coastal monitoring and management strategies; but as non-specialists, ocean colour end-users often do not have the necessary knowledge regarding algorithm limitations and uncertainties in order to make informed choices about the standard available chlorophyll products. Results from chapter 3 have demonstrated that it is unlikely that any one available satellite algorithm can successfully retrieve [Chl *a*] information with an acceptable degree of uncertainty over all the in-water conditions found in the South African coastal region. OWT classification facilitates the water type-specific application and blending of the most appropriate algorithms on a per-pixel basis, providing greatly improved capability to provide a single blended product that may be used across a wide range of water types and systems.

4.1.4 Chapter aims

This chapter builds upon the OWT characterization and validation performed in chapter 3 by demonstrating the broader application potential of the newly defined OWTs through classification of the entire archive of Level 2 data from MERIS; the availability of the 709 nm band made MERIS the optimal choice for OWT classification applications since the current study's scope included productive phytoplankton-dominated waters. This chapter further aims to demonstrate the use of these OWTs in assessing the OWT phenology and bio-optical variability, in addition to its potential impacts on realised usability of ocean colour satellite products, in the waters of South Africa. The utility of the classification and algorithm blending framework for retrieving opimized [Chl a] across dynamic concentration ranges and different water types, is assessed in comparison to standard and regional satellite products. Finally, the overall classification capability for South African coastal waters is discussed in terms of the constraints introduced by the current set of OWTs, satellite temporal coverage, and reflectance product quality.

4.2 Methods

4.2.1 Classification and algorithm blending procedure as applied to MERIS data

The eleven optical water type classes (figure 4.2) as defined in chapter 3 were used to classify ten years of reduced resolution (1 km) MERIS data for South African coastal waters. For simplification the coastal zone of the country was divided into three domains as shown by the bounding boxes in figure 2.3; these domains were further defined as the west coast (30 to 35 °S, 16 to 19 °E), the south coast (33 to 37 °S, 18 to 28 °E) and the east coast (27 to 36 °S, 27 to 33 °E). The MERIS archive from May 2002 to April 2012 was searched for any swaths that contained pixels within the predefined boxes, which were subsequently extracted. Regional subsets included the water-leaving reflectance bands, level 2 flags and algal pigment indexes 1 and 2; these data were added to the time series of each respective region. For input to the classification procedure the MERIS water-leaving reflectance was converted to remote sensing reflectance (R_{rs}) according to equation 2.2 (Antoine and Morel, 2005). Only data from MERIS bands 2 to 10 (centred at 442.5, 490, 510, 560, 620, 665, 681.25, 708.75 and 753.75 nm) were used during classification; band 1 (412.5 nm) was not used, as in previous studies (Vantrepotte et al., 2012; Moore et al., 2014), since this band in MERIS data from the 3rd reprocessing is generally noisy (Smith et al., 2013; Zibordi et al., 2013; Cristina et al., 2014).

The classification and algorithm blending procedure is detailed in sections 2.3.2 and 2.3.4, and entails the following:

- The degree of closeness of the satellite R_{rs} spectra to each OWT is quantified by means of the Mahalanobis distance; this step requires the mean spectrum and covariance matrix of each OWT
- Class membership functions of each pixel to each OWT is calculated by using the Mahalanobis distance value as input to a probability density function
- Satellite pixels that were flagged for uncorrected glint and cloud were excluded from classification
- The dominant class per pixel is the OWT to which the pixel shows the highest membership; this is used to create the dominant class maps

- Class membership functions were used to weight and blend class-appropriate algorithms (as defined in section 3.2.6) on a per-pixel basis
- Satellite pixels that were flagged for reflectance quality (PCD1_13) were excluded from the algorithm blending process and resulting blended [Chl *a*] product

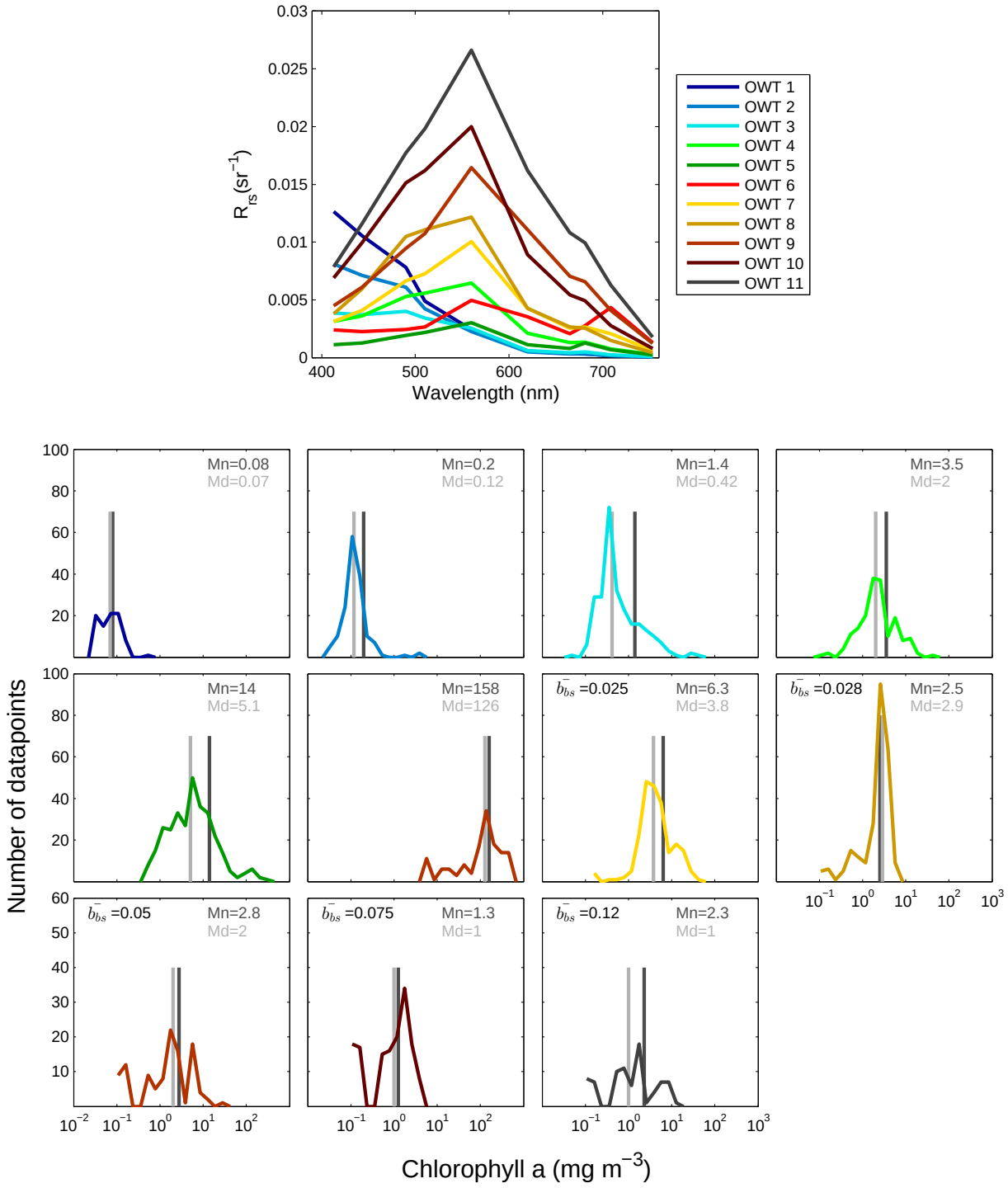


Figure 4.2: The mean R_{rs} spectra of the eleven optical water type classes that were used to classify MERIS reflectance data of South African coastal waters, as well as the corresponding [Chl a] ranges and frequency of occurrence. Also included are the class means (dark grey) and medians (light grey), as well as the mean $b_{bs}(560)$ for the elevated reflectance classes.

4.2.2 Optical water type phenology and analysis: regridding and metric design

In order to assess the spatial and temporal persistence of the OWTs in the coastal regions for South Africa the daily class memberships, chlorophyll products, and relevant level 2 flags were regridded onto a common 1/100° grid using linear interpolation; this was done for the period of the entire MERIS mission. Data were only regridded after OWT classification to enable classification and algorithm application on unaltered R_{rs} spectra.

The frequency of dominant OWT was performed according to the methods used in Moore et al. (2014); if the membership sum of a pixel exceeded a threshold value of 0.10, the OWT with maximum membership for that pixel was counted. The same threshold was used to define regions of low membership, which were counted each time the total membership sum of a pixel was below 0.10. These counts were normalized by the total count of valid water pixels (not affected by cloud or uncorrected sun glint) over the entire MERIS mission. The threshold value of 0.10 was adopted from previous studies using the Mahalanobis distance for the classification of satellite ocean colour pixels (Mélin et al., 2011; Vantrepotte et al., 2012; Moore et al., 2014). The value is however arbitrary and could be adjusted according to the study needs; for example the elevated reflectance clusters generally classify less well due to their relatively large reflectance; spectra with similar shapes having different amplitudes might have a much larger Mahalanobis distance (and thus smaller membership) compared to moderate reflectance classes.

The utility of a satellite mission in the dynamic coastal environment is critically dependent on the realised revisit time, i.e. how frequently usable earth observation data are available for a given ecosystem. For this reason, the frequency of occurrence of the cloud, uncorrected glint, and the MERIS reflectance confidence (PCD1_13) flags, normalized by the total number of coverage pixels, were assessed. Where applicable, frequency measurements were arranged according to austral seasons, i.e. winter was represented by all data from June, July and August, spring by September, October, November, and so forth.

4.2.3 Case studies demonstrating OWT classification and algorithm application

Several case studies were selected to demonstrate the utility of the eleven OWTs for MERIS image classification and Chl *a* algorithm blending in each of the coastal regions. These examples included two high biomass algal bloom events to demonstrate the dynamic range of coverage of the classification and algorithm blending framework, as well as a flood which demonstrates the classification and algorithm application in a short-lived Case 2 scenario. Each study shows the dominant OWT map (also known as the hard classification map), the total membership sum map, maps for the relative weighted contribution of each of the five Chl *a* algorithms, and the final blended [Chl *a*] product.

In order to better understand the flood event example, daily flow rate data were obtained for the Thukela River (location shown in figure 4.30), the third-largest river in southern Africa (Bosman et al., 2007), from the website of the Department of Water and Sanitation (DWA, 2016). Data were acquired for the period of 1 May 2002 to 30 April 2014 at the Mandini station (V5H002) as it represents the closest approximation of the flow conditions of the Thukela estuary; the station is located approximately 23 km upstream of the river mouth.

Medium-scale shapefiles for South African rivers were obtained from the Resource Quality Information Services website (RQIS, 2016); these data were only utilized in the case studies for the maps of the elevated reflectance classes, in order to demonstrate the large quantity of rivers contributing to the coastal environment. Shapefiles from the CIA World DataBank II (Central Intelligence Agency, 1977) were used in all other maps to display only the relatively large rivers of South Africa.

4.2.4 Chl *a* algorithm time-series analysis

The Chl *a* algorithm blending procedure (as outlined in section 2.3.4 which uses equations 2.20 and 2.21) was applied to the MERIS reflectance data for the entire mission period. Pixels flagged for uncorrected sun glint and/or PCD1_13 were replaced with null values and were resultantly not used in calculations. Within each of the three primary coastal regions subregions were defined with the purpose of assessing the blending algorithm performance within varying trophic and bio-optical regimes. The subregional averages for the blended product were compared to those of the standard MERIS Chl *a* products, in addition to regional Chl *a* products (if available). For each product the daily average, as well as a 30-day running average, were assessed.

In order to assess the performance of the blended Chl *a* product in South African coastal waters the blended product was compared to other available regional and standard MERIS products with spatially averaged time series of [Chl *a*] for a selection of subregions. MERIS products included the standard Level 2 Case 1 and Case 2 products, Algal1 and Algal2 respectively. In the southern Benguela regional products included [Chl *a*] from the EAP and MPH algorithms. The EAP is a semi-analytical inversion algorithm specifically designed for the southern Benguela; more details on the EAP can be found in section 2.3.1, as well as in Evers-King et al. (2014) and Evers-King (2014). The MPH is a red/NIR baseline subtraction algorithm applied to MERIS bottom-of-Rayleigh reflectances which was designed for high biomass coastal and inland waters of South Africa; more details on the MPH algorithm can be found in Matthews et al. (2012). Data for both the EAP and MPH were available from October 2002 to April 2012.

4.2.5 MERIS product limitations

This section aims to serve as a diagnostic tool and user guide for ocean colour utility by providing realised usable ocean product observation frequency for the region. This was achieved by assessing the seasonal persistence of three MERIS flags that were used to maintain the quality of data during the OWT persistence and [Chl a] time series studies, including pixels that were flagged as cloud, uncorrected sun glint, or for insufficient quality of the reflectance product. The combined effect of these flags as they pertain to the valid pixels for time series analysis were also assessed. This has important implications for ocean colour users, which will be discussed in section 4.4.3.

Persistence of the cloud flag

The presence of cloud over South Africa displays regional and seasonal variability (figure 4.3). The highest persistence of clear skies (about 20% cloud) can be found to the north-west of Cape Town during summer when south-easterly winds prevail over the region. There is a band of increased persistence of the cloud flag next to the coast north of St Helena bay in summer (>30%) and autumn (>45%) which coincides with the west coast fog season. Olivier (2002) showed that the region between Cape Columbine and Port Nolloth (27.5 °S) experienced high to medium intensity fog with the highest frequency of occurrence during autumn. Increased cloud persistence along the east coast and the off-shore ocean during spring and summer corresponds to the presence of tropical-temperate troughs; these synoptic-scale systems form northwest to southeast bands of rainfall (and cloud) over Southern Africa and the Southwest Indian Ocean, and are responsible for the bulk of South African summer rainfall (Harrison, 1984; Washington and Todd, 1999). There is also high persistence of cloud over the Agulhas Current during spring and summer; studies have noted the presence and persistence of cumulus clouds over the Agulhas Current to the east of 20 °E, usually with conditions of >20 °C and persistent along-current winds (Lutjeharms et al., 1986; Lee-Thorp et al., 1998; Lutjeharms and Rouault, 2000). The high persistence of cloud to the south of the country during winter is most likely due to equator-ward shift of the mid-latitude cyclones during austral winter.

Sub-pixel heterogeneity - in this case the presence of both cloud and water within a pixel - has been a noted problem for the MERIS atmospheric correction process (Brockmann, 2006). Although the MERIS 3rd data reprocessing has improved the cloud identification over water, particularly at the borders of clouds and less bright scattered clouds (Lerebourg and Bruniquel, 2011), there may still be a problem with cloud shadows; Brockmann

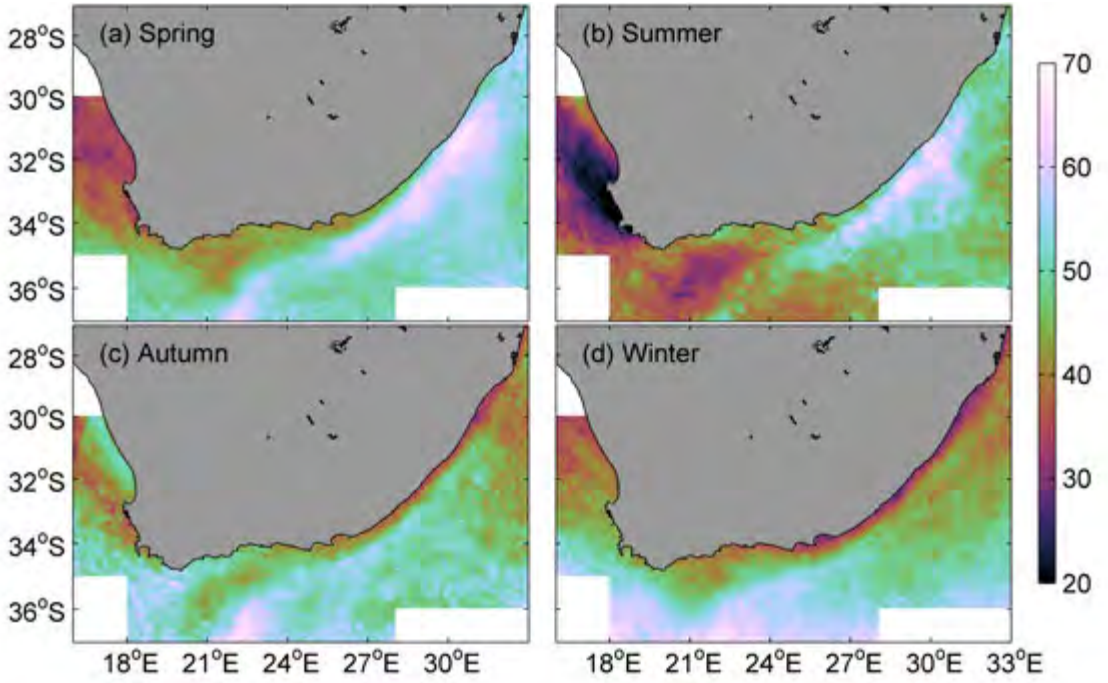


Figure 4.3: The seasonal persistence (%) of cloud for MERIS data from 2002 through 2012 in South African coastal waters

(2006) demonstrated how reflectance spectra became increasingly negative close to clouds. Within the OWT framework this may lead to unclassified pixels, or cause the pixel to be incorrectly classified as a more absorbing (or higher biomass) OWT, which could possibly result in the application of an inappropriate algorithm and/or an overestimation of [Chl *a*].

Persistence of the uncorrected sun glint flag

Sun glint is the component of specular reflection of light from the surface of the water body. It is a function of the sun position, sensor viewing angle, and the state of the sea surface. Since the signal from sun glint may be much larger than the water-leaving radiance signal, the affected regions show erroneously high reflectance and are generally flagged or masked out in satellite data products. Although some sensors are capable of tilting the sensor away from nadir in order to avoid contaminated areas (e.g. SeaWiFS), MERIS has a fixed viewing point at nadir which results in the exclusion of up to 50% of the spatial coverage at subtropical latitudes due to sun glint (Steinmetz et al., 2011); this effect can be seen in the southern Benguela (figure 4.4) during summer with >40% persistence of uncorrected sun glint. The highest persistence of the flag occurs during summer and spring and corresponds to the regions of lowest cloud persistence during these periods.

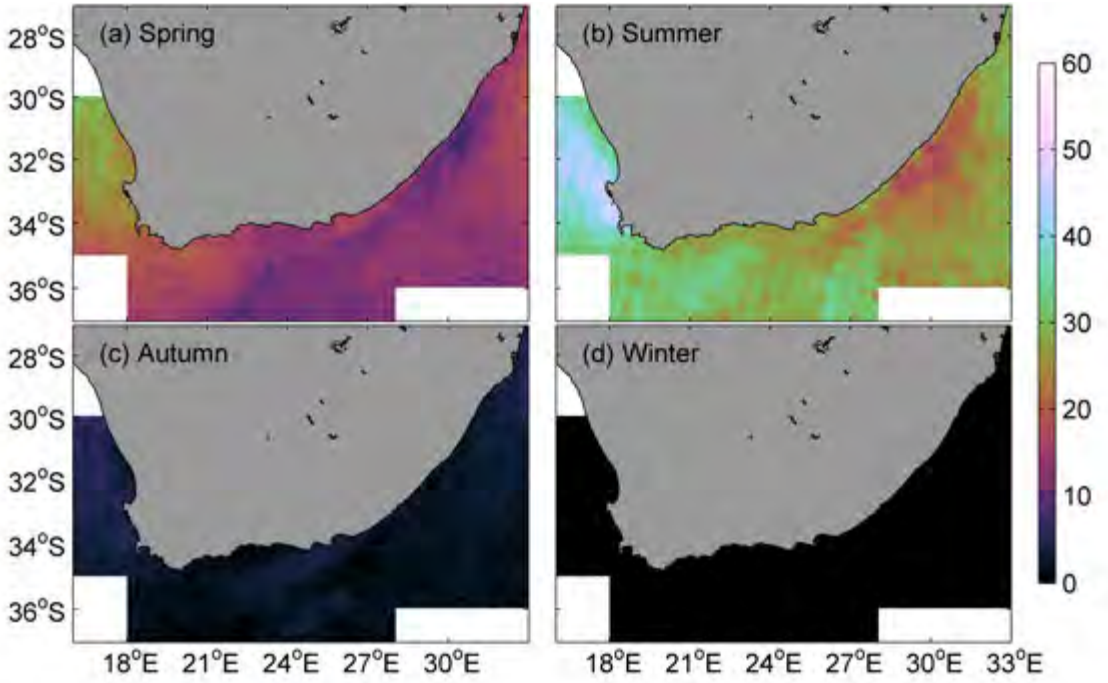


Figure 4.4: The seasonal persistence (%) of the uncorrected sun glint (higlint) flag for MERIS data from 2002 through 2012 in South African coastal waters

Due to the seasonal change of sun angle relative to sensor position the flag displays much lower occurrence (<10%) in all regions during autumn and winter. Although algorithms exist that correct for or minimizes the effects of sun glint (see references in Kay et al., 2009), there is still concern about uncertainty in glint affected waters. It is generally recommended that future sensors have a tilting mechanism to avoid sun glint (McClain et al., 2014); the MERIS follow-on mission OLCI has a field-of-view that is tilted to reduce sun glint pollution (Lavender and Kay, 2010).

Persistence of the reflectance quality flag

The PCD1_13 is the MERIS confidence flag for the thirteen MERIS water-leaving reflectances and indicates that the atmospheric correction has failed in at least one of the bands (Brockmann, 2006); it may be triggered by the ice-haze and/or out-of-aerosol database flags, due to uncorrected glint and/or low sun angles. The 3rd MERIS reprocessing introduced a negative tolerance threshold performed on TOA water reflectance that is less exclusive to small amounts of negative reflectance at various wavelengths which could be attributed to sensor noise (Lerebourg and Bruniquel, 2011). Even though some of the standard MERIS chlorophyll algorithms may still function if the flag is raised, the reflectance classification and blending method could be negatively affected by higher reflectance

uncertainties.

Figure 4.5 shows that this flag has high seasonal variability. The flag generally shows lower persistence during winter and autumn, whilst it is raised >50% of the time during summer; the flag is also triggered by the uncorrected sun glint flag, which would explain the increased persistence of the PCD1_13 flag during summer and spring. The flag was often raised in the vicinity of cloud edges and shadows and resultantly has a degree of spatial overlap with the persistence of cloud. The near-shore zone of the east coast is particularly affected, with the flag persisting >70% in all but the winter months. Inspection of these flagged reflectances (not shown) revealed negative reflectance in the blue and red for absorbing spectral shapes (like OWTs 4 and 5), and negative reflectances in the red for oligotrophic water shapes (like OWTs 1 and 2), suggesting persistent problems with aerosol correction; radiometric noise and calibration errors in the NIR can be propagated through the atmospheric correction to contribute to errors of the retrieved marine reflectances (Jolivet et al., 2007).

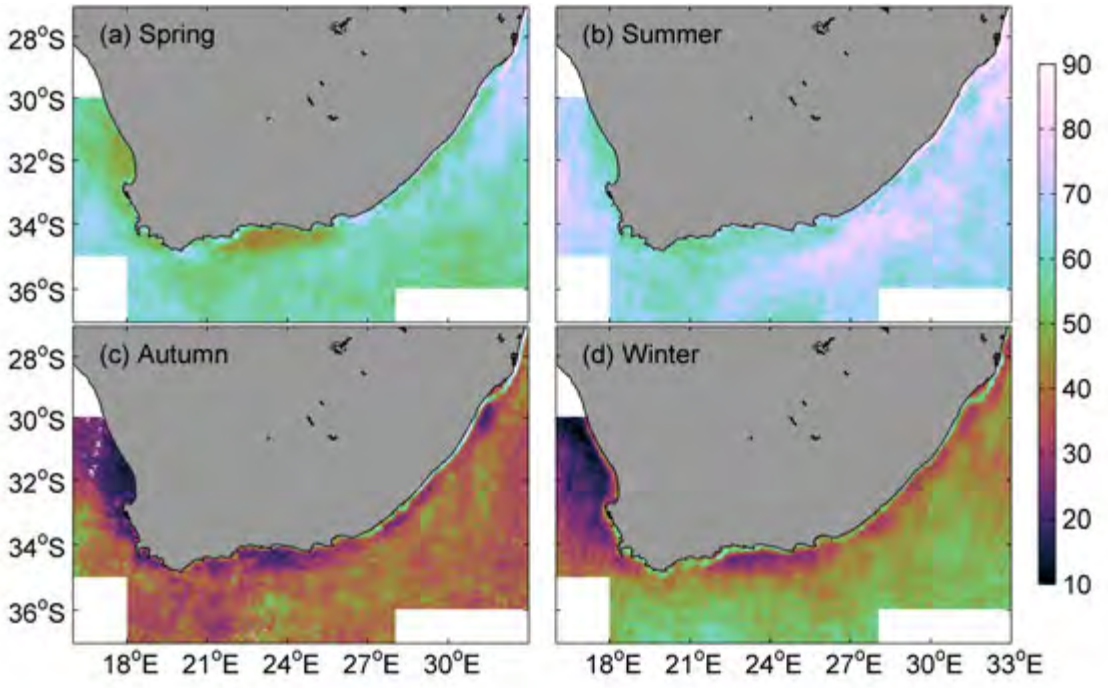


Figure 4.5: The seasonal persistence (%) of the reflectance quality flag (PCD1_13) for MERIS data from 2002 through 2012 in South African coastal waters

Realised MERIS coverage and valid pixels

In order to assess the potential variability in regional satellite data coverage the total footprint of MERIS is compared to the actual valid pixels that are available for use after scientific and data quality flagging; the comparison was done for the period of 2002 though 2012 and is shown in figure 4.6. The data were presented seasonally since there was clear seasonal variability, while two different variants of *valid pixels* were used for normalization of time series data. The pixels that were not affected by cloud or uncorrected glint were used in the OWT persistence calculations in order to test the classification performance on maximum returns. However, even though the classification algorithm was often able to classify irregular spectra, the performance of the blended chlorophyll product was generally compromised; resultantly the blending algorithm was only applied to pixels that were unaffected by the PCD1_13, cloud or uncorrected sun glint flags. The Agulhas Current regions showed generally less valid pixels than surrounding waters, which was also the case for Mélin and Vantrepotte (2015) when using SeaWiFS MLAC data, as well as for Krug et al. (2014a) when using various Infra-Red sea surface temperature data products. The least valid pixels are available during summer and spring for both variants, probably due to the high persistence of fog, cloud, uncorrected glint, and difficulties with the atmospheric correction during these months (as demonstrated in section 4.2.5).

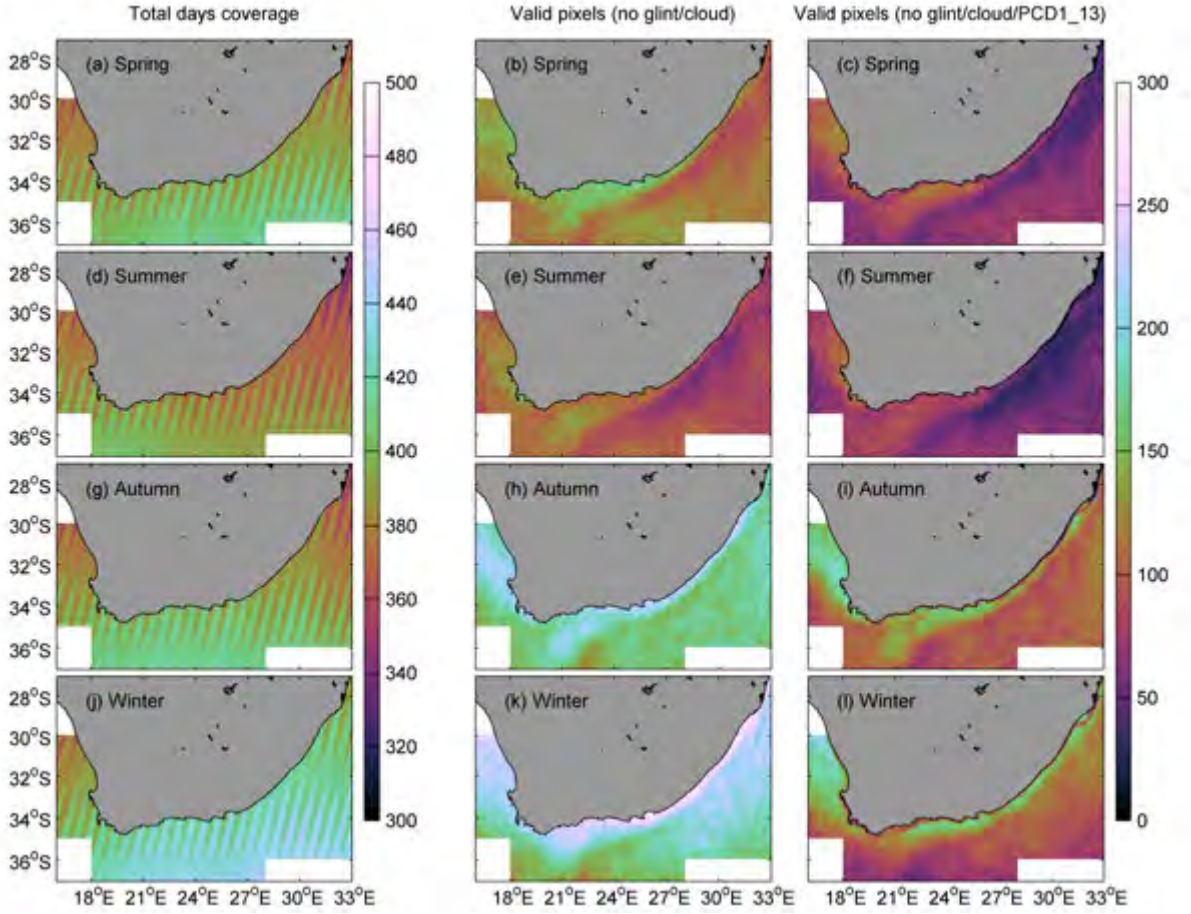


Figure 4.6: The left-hand column shows the number of days per pixel with MERIS coverage per season; the middle column shows the corresponding number of pixels that were not affected by uncorrected sun glint or cloud per season (used for OWT persistence calculations); whilst the right-hand column shows the number of pixels that were not affected by uncorrected sun glint, cloud or the PCD1_13 flag per season (used for blended chlorophyll product calculations). All data are calculated from the entire MERIS mission from 2002 through 2012. Note that the left column has a separate colour bar.

4.3 Results

4.3.1 Shelf-scale persistence of the optical water types in SA coastal waters

This section provides maps of the dominance and persistence of the individual OWTs over the total MERIS mission from 2002 to 2012. Maps of the dominant and second most dominant OWTs (per pixel) are given in figure 4.7. The percentage frequency of dominance of the optical water types 1 through 5, as well as OWT 7, are shown in figure 4.8. OWTs 6 and 8 though 11 were only presented and discussed at a regional and sub-regional scale, due to the near-coastal extent and generally small-scale occurrence.

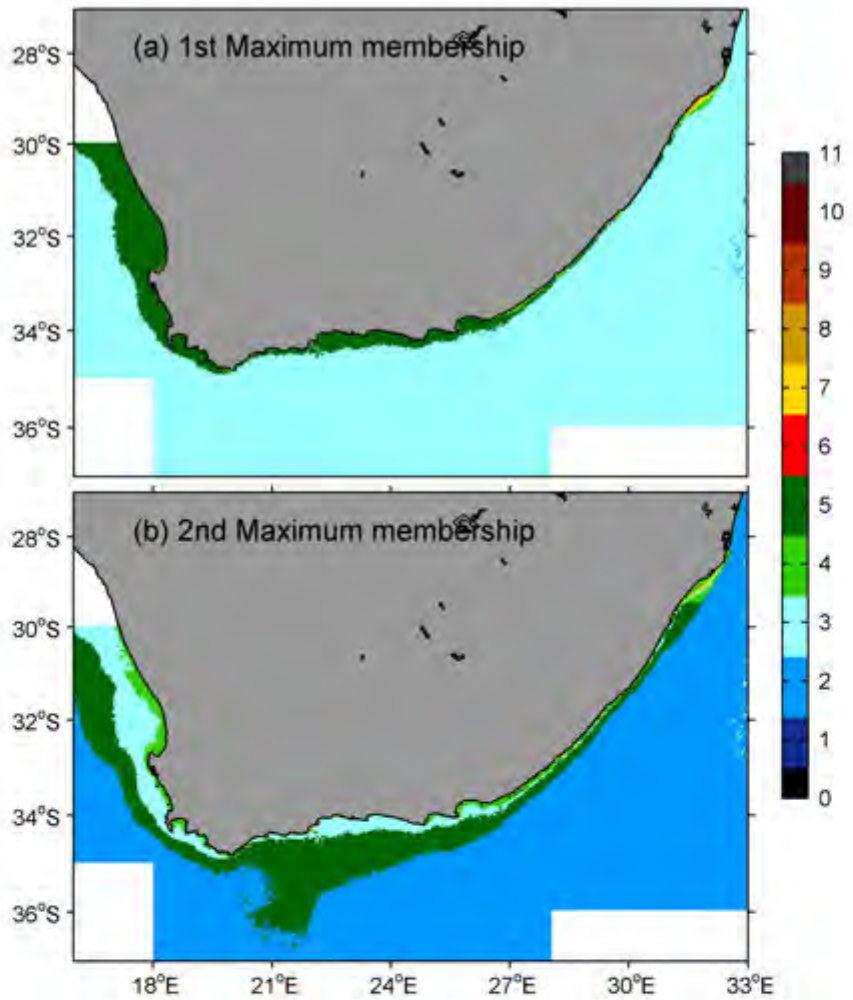


Figure 4.7: The hard classification map of (a) the dominant optical water type, as well as (b) the second most dominant optical water type per pixel in the coastal waters of South Africa for MERIS data from 2002 through 2012.

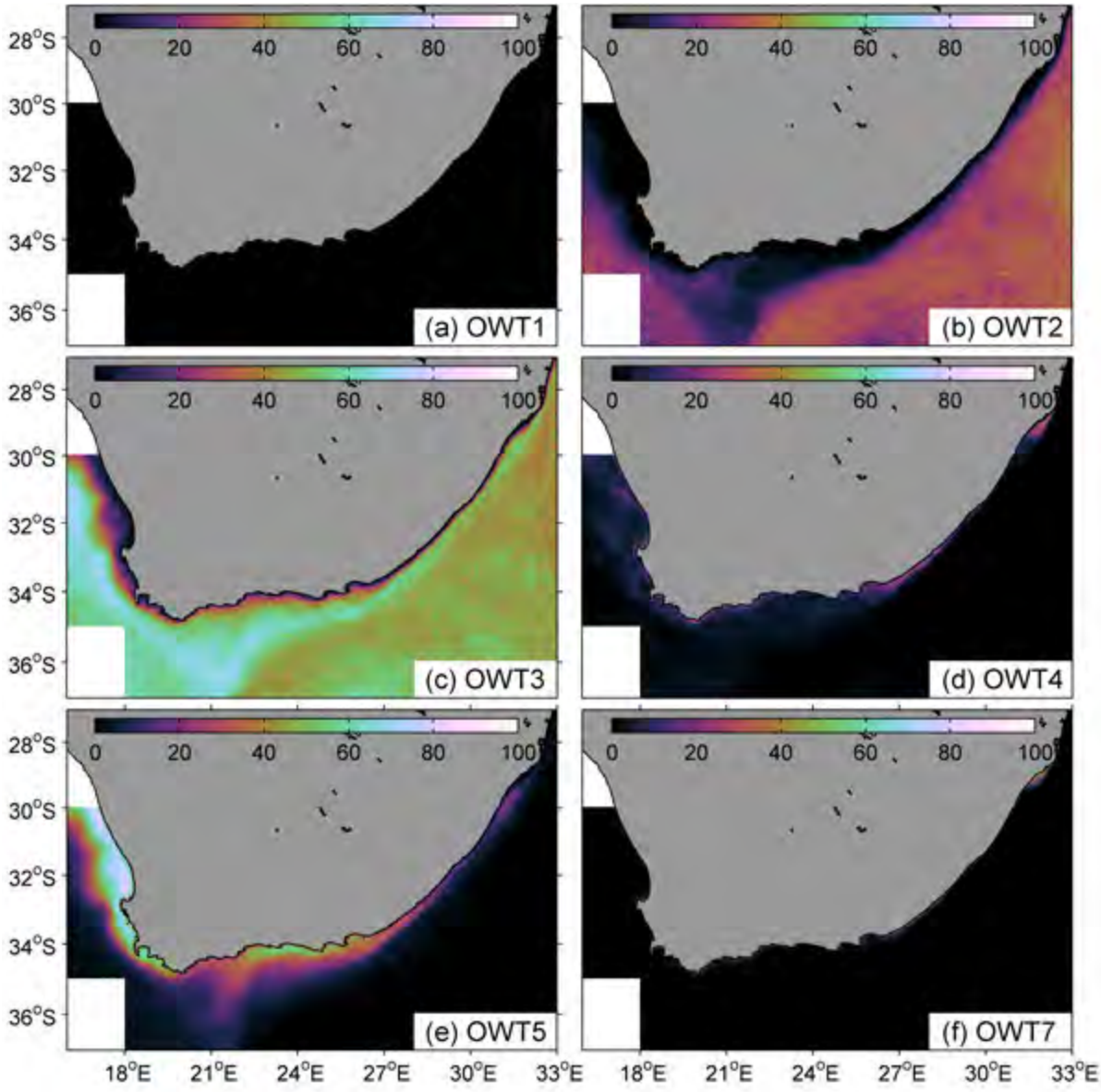


Figure 4.8: The percentage dominance of (a) OWT1, (b) OWT2, (c) OWT3, (d) OWT4, (e) OWT5, and (g) OWT7 in the coastal waters of South Africa for MERIS data from 2002 through 2012.

The west coast

The high biomass class, OWT5, dominates on the shelf along the west coast approximately 70% of the time (figure 4.7a and 4.8e). This class shows synoptic coherence to the underlying bathymetry and could potentially be related to the strength and position of the shelf-break front (Shannon and Nelson, 1996); the offshore extent of OWT5 is given in the secondary dominance map (figure 4.7b), representing the sporadic offshore displacement of these high biomass waters during strong upwelling. OWTs 3 and 4 can also be found, to a

much lesser extent, on the shelf and in the near-shore zone respectively; the occurrence of these classes on the shelf may be associated with the variability in phytoplankton biomass induced by the pulsed upwelling system. The low dominance of OWT4 may suggest that it is a short-lived or more dynamic water type; this will be discussed further below. The dominance and persistence of individual OWTs in this region will be discussed in more detail in section 4.3.3. The offshore zone is dominated by OWT3; this class represents reflectance data associated with low to moderate [Chl *a*], which, at the continental scale, encompasses regions ranging between offshore oligotrophic and mesotrophic shelf waters. OWT2, associated with oligotrophic oceanic waters, shows secondary dominance in the offshore regions.

The south coast and Agulhas Bank

OWT3 appears to be the primary water type for most South African shelf and offshore regions; this class shows the highest dominance over the Agulhas Bank (figure 4.8c) and is, in this case, indicative of the mesotrophic shelf sea environment. The high biomass class, OWT5, dominates in the coastal regions along the south coast that are affected by wind-induced upwelling; these findings are consistent with previous studies that also noted elevated phytoplankton biomass in these coastal regions (Shannon et al., 1984; Demarcq et al., 2003; Weeks et al., 2006; Demarcq et al., 2007). The dominance and persistence of individual OWTs in this region will be discussed in more detail in section 4.3.4. OWT5 also occasionally extends over the Agulhas Bank on the inner edge of the Agulhas Current, visible in both figures figure 4.7b and 4.8e, possibly resulting from phytoplankton being entrained at the coast and moved offshore by the current (Jackson et al., 2012), or due to nutrient availability from the 'cold ridge' extending into this region (Probyn et al., 1994; Demarcq et al., 2003). The offshore regions show secondary dominance of OWT2, representative of low biomass oligotrophic waters; the western Agulhas Bank appear to have a slightly higher persistence of this class than the eastern side, consistent with previous reports of the strong oceanic forcing of this region (Largier et al., 1992). Occasional dominance of OWT2 off the western Agulhas Bank and offshore from the west coast may be associated with the passage of warm nutrient-poor Agulhas Rings (Garzoli and Gordon, 1996; Schonten et al., 2000), eddies that are shed from the Agulhas Current retroflection.

The east coast

Unlike the other coastal regions, OWT3 dominates very close to shore along the east coast, with slightly elevated persistence over the Agulhas Current (figure 4.8c). There is a more homogeneous distribution between the persistence of the low biomass OWT2 (figure 4.8 b) and OWT3 to the east of the Agulhas Current, indicative of the characteristic near-oligotrophic nature of this offshore region. The very low biomass class, OWT1 (figure 4.8a), displays very low dominance (generally less than 2%, not visible at this colour scale) and is not often selected as the dominant class; there is the possibility that the class is not parameterized optimally for MERIS classification since it was primarily comprised of synthetic data. OWT4 shows dominance in a thin coastal strip to the east of Algoa Bay, as well as in the northern section of the KwaZulu-Natal (KZN) Bight; both regions are subject to inshore-edge upwelling (Schumann, 1987; Lutjeharms et al., 2000a; Goschen et al., 2012) and divergence-driven upwelling (Lutjeharms, 2006) when the Agulhas Current follows the shelf break and moves away from the shore. OWT5 also dominates intermittently in this thin coastal strip and shows secondary dominance over the southern KZN Bight and may be related to occasional enhanced phytoplankton productivity from fluvial nutrient input. OWT7 (figure 4.8f), representing moderate biomass in moderately scattering waters, is the only elevated reflectance class that is visible at the continental scale. This class dominates the near-coastal zone in the northern KZN Bight, where fluvial suspended sediments contribute to the turbidity of the surface waters. OWT5 shows secondary dominance over the southern shelf of the KZN Bight, but is only dominant about ∼20% of the time. The dominance and persistence of individual OWTs in this region will be discussed in more detail in section 4.3.5.

4.3.2 Shelf-scale representativeness of the optical water types in SA coastal waters

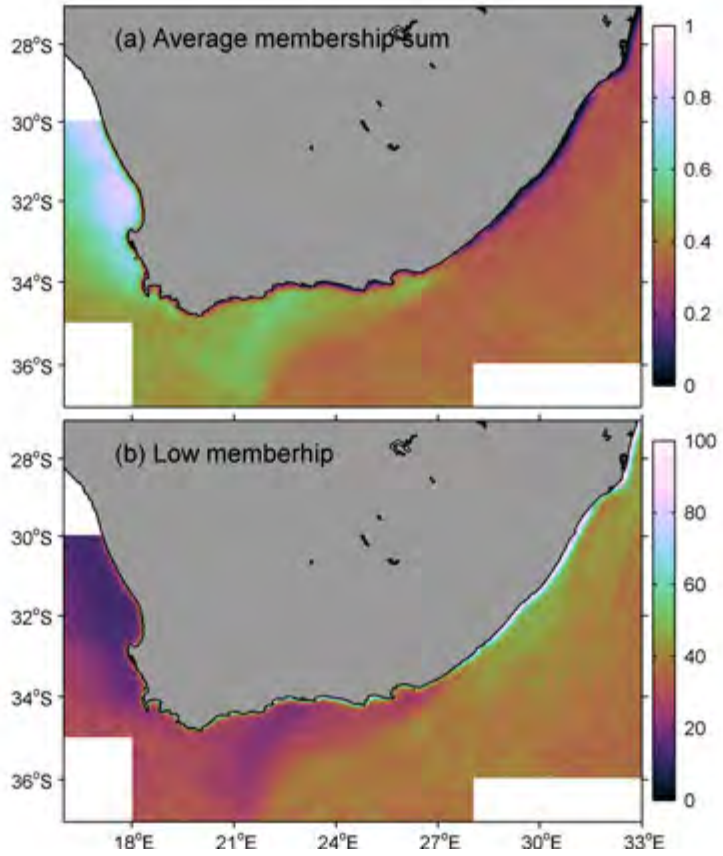


Figure 4.9: The (a) average membership sum and the (b) percentage low membership (total membership sum < 0.10) in the coastal waters of South Africa for MERIS data from 2002 through 2012.

The ability of the current set of OWTs to resolve the optical variability in SA coastal waters was assessed through the persistence of the average membership sum (figure 4.9a) and low membership (figure 4.9b), which displayed an inverse relationship. Mélin and Vantrepotte (2015) used a domain average of the total class membership as the measure of satisfactory classification, where the desired result is around 1 (although it was higher at 1.14 for their study). However, the SA domain average of the total class membership was only 0.41, broken down as 0.55, 0.41, and 0.33 for the west, south and east coast domains respectively. The classification performance was highest along the west coast, which is to be expected since the OWTs were primarily developed with data from the southern Benguela. The highest average membership can be found on the continental shelf off the west coast (>0.6), whilst membership sums decrease slightly offshore and towards the south-east over the Agulhas Bank (0.4 - 0.5); even lower sums are seen over the Agulhas

Current and Indian Ocean. The lowest membership sums persist in the near-shore zones along the south coast and a strip of very low membership sums (<0.1) in the coastal zone east of approximately 28°E (<0.1); this is only interrupted in the northern KZN Bight where membership sums increase slightly to about 0.3. The map of the persistence of low membership maps, shown in figure 4.9b, shows similar findings to the membership sum maps, where the highest percentage occurrence of low membership sums (>90%) in the near-shore zone on the east coast, with a similar interruption in the northern KZN Bight. Low memberships usually occur between 25 and 40% of the time over the Indian Ocean, whilst persistence may be as high as 50% during autumn (not shown) over some of the Agulhas Current. To the west of the Agulhas Current the persistence of low memberships are generally limited to below 30%, with the lowest persistence of low membership over most of the west coast shelf regions. Unclassified pixels and low membership sums have often been attributed to difficulty with the atmospheric correction process and missing optical water types (Vantrepotte et al., 2012; Moore et al., 2014; Mélin and Vantrepotte, 2015). The high persistence (>75%) of the PCD1_13 flag (figure 4.5) in the near-shore waters along the east coast may indicate that the low persistence in this area could very likely be attributed to atmospheric correction problems.

4.3.3 Case study 1: St Helena Bay, southern Benguela

In the southern Benguela, the St Helena Bay region is a known retention zone (Lett et al., 2006) with frequent occurrence of elevated phytoplankton concentrations (Weeks et al., 2006; Demarcq et al., 2007); these trends are visible in the persistence of the current set of OWTs where the continental shelf is dominated by the moderate reflectance classes 4 and 5 (figure 4.8), which represent the medium to high biomass signals, whilst the elevated reflectance OWTs persist in the near-shore zone as demonstrated by figure 4.10. These elevated reflectance classes show the highest persistence along the south-eastern shores of St Helena Bay, which is potentially attributable to re-suspension of shallow bottom sediments through wave action, or possibly due to visibility of the shallow bottom (<15 m in this area) on days with low levels of turbidity. Although water clarity at the Berg River mouth (location shown in figure 4.16) varies seasonally with river discharge (Schumann, 2009), the elevated reflectance classes did not show seasonal variability, which suggests that these classes are not well parameterized to detect different types of fluvial sediment loads.

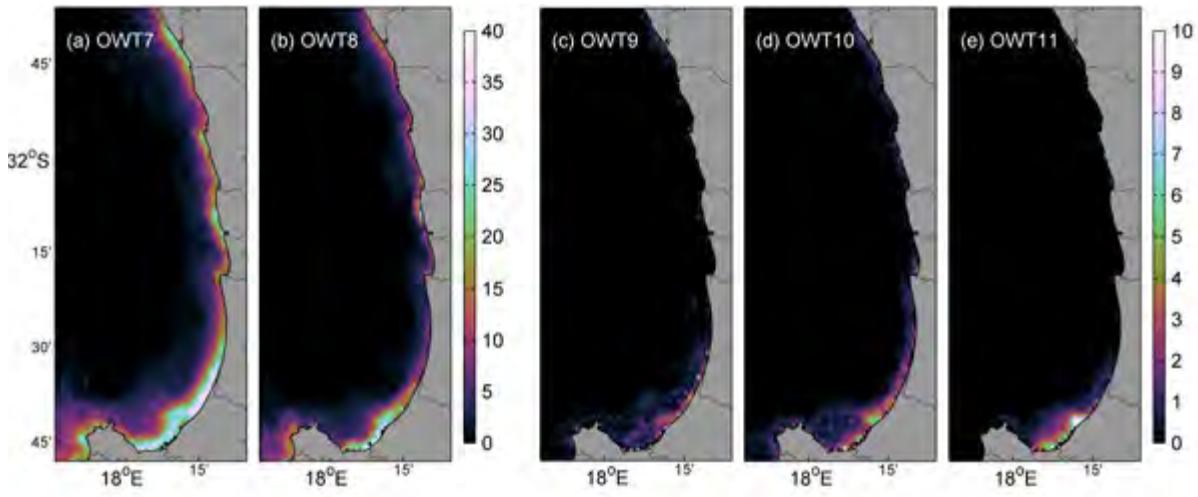


Figure 4.10: The percentage dominance of (a) OWT7, (b) OWT8, (c) OWT9, (d) OWT10 and (e) OWT 11 in St Helena Bay for MERIS data from 2002 through 2012. The dark grey lines represent all the rivers in the region.

There appears to be some seasonal variability in the moderate reflectance OWTs (figure 4.11), which could potentially be linked to the strength of the upwelling and offshore Ekman transport in the southern Benguela. Whilst recently upwelled water has low [Chl *a*] (<1 mg m^{-3}), maturing upwelled water and aged water both have higher ranges of 1-20 mg m^{-3} Chl *a* and 5-30 mg m^{-3} Chl *a* respectively (Barlow, 1982; Barlow et al., 2005); OWT4 could potentially represent the early stage of maturing upwelled water with

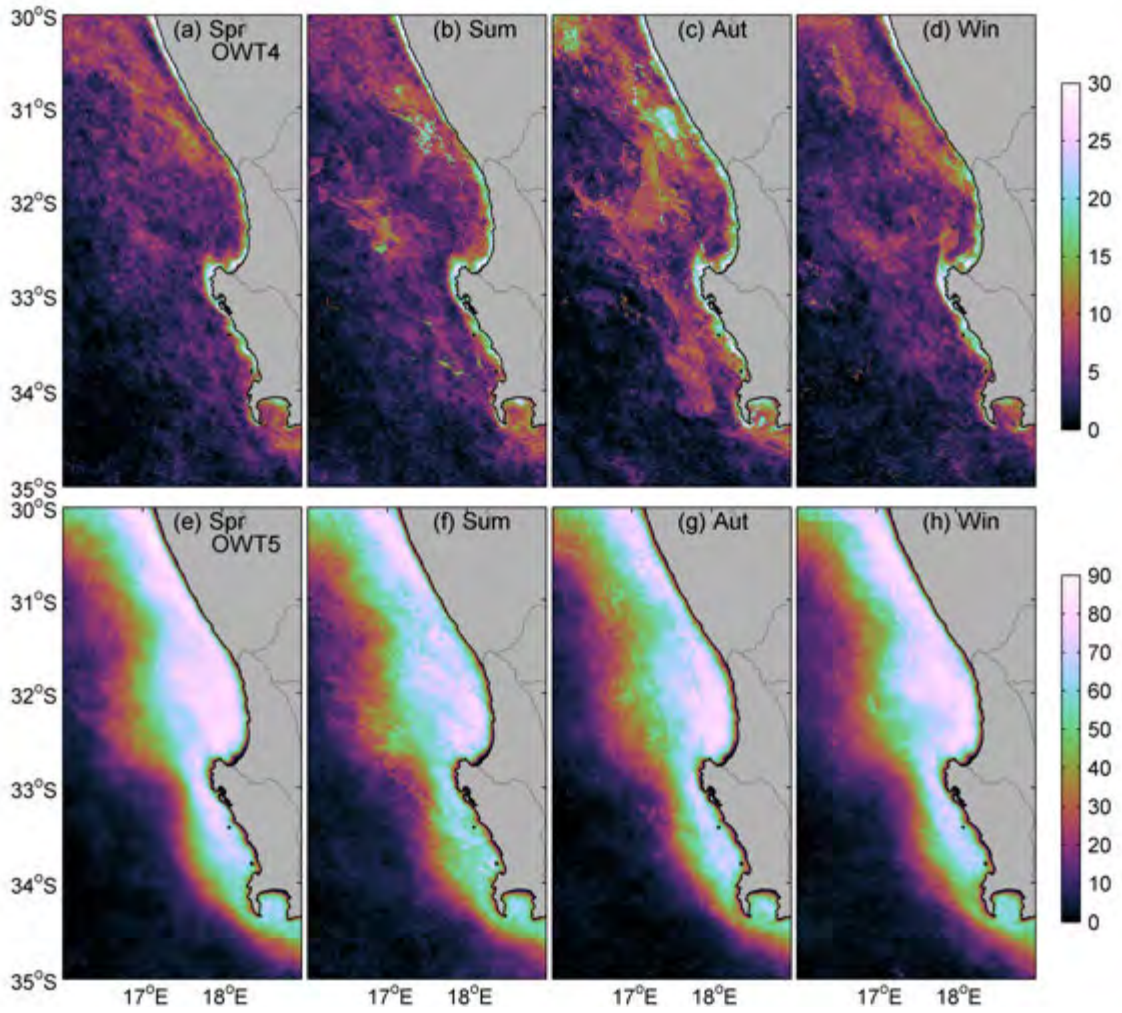


Figure 4.11: The percentage seasonal dominance of OWT4 (panels a - d) and OWT5 (panels e - h) along the west coast of South Africa for MERIS data from 2002 through 2012. Note the change of scale between the two water types.

moderate [Chl *a*], whilst OWT5 may represent aged water with higher [Chl *a*]. Upwelling-favourable winds tend to peak in summer (Pitcher et al., 2010), whilst Ekman transport is likely to be strongest where the continental shelf is narrowest (Nelson and Hutchings, 1983). These conditions may play a role in the less frequent occurrence of OWT5 between Cape Point and Cape Columbine during summer months (figure 4.11f), when regions of stronger upwelling may display lower [Chl *a*] relative to surrounding coastal areas (Weeks et al., 2006). Conversely, OWT4 shows slight elevation in dominance near 31°S during summer and autumn (figure 4.11), which could potentially be related to lower [Chl *a*] associated with newer upwelled water. The patchy persistence of OWT4 in the offshore regions of the southern Benguela may result from offshore transport during intense active upwelling (Brown et al., 1991; Weeks et al., 2006), signifying the dynamic nature of this

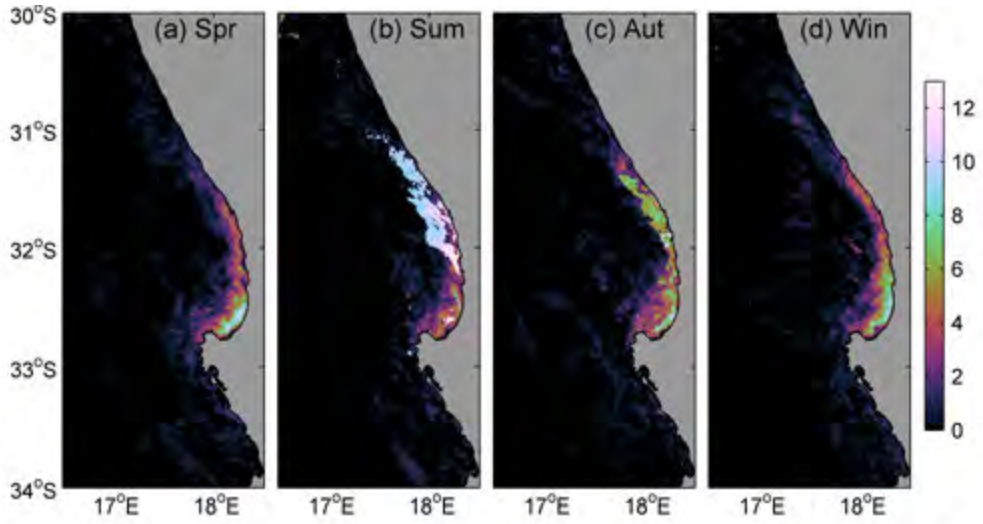


Figure 4.12: The percentage seasonal dominance of OWT6 along the west coast of South Africa for MERIS data from 2002 through 2012.

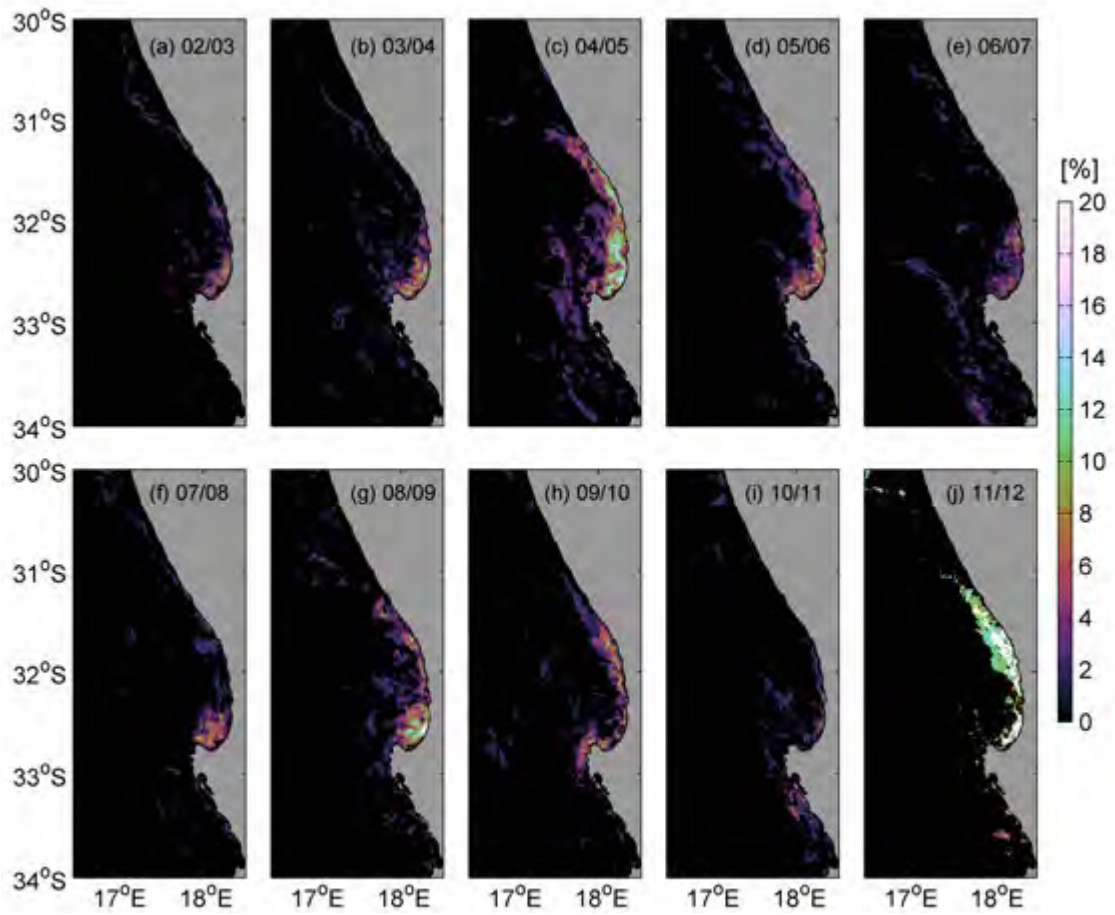


Figure 4.13: The percentage dominance of OWT6 for each annual upwelling season between 2002 and 2012 along the west coast of South Africa for MERIS data. Note that the upwelling seasons are between October and May.

class along the west coast.

Community structure in the southern Benguela has been shown to be primarily dominated by diatoms (Lamont et al., 2014; Barlow et al., 2016; Van der Lingen et al., 2016) particularly in the inshore zone (Barlow et al., 2005) where these species are able to take advantage of the nutrient-rich and turbulent upwelling environment; however, a wide variety of functional types may dominate at any stage depending on the stratification and upwelling state (Crichton et al., 2013). The Benguela periodically suffers from Harmful Algal Blooms (HABs) which are usually attributed to dinoflagellate species (Pitcher and Weeks, 2006). These blooms occur most frequently from January to May in the latter half of the upwelling season (Pitcher and Calder, 2000) associated with increasingly stratified conditions (Pitcher and Weeks, 2006). HABs can negatively impact commercial and recreational activities in the region (Pitcher and Cockcroft, 1998; Pitcher and Calder, 2000; Probyn et al., 2000). Within the current classification framework OWT6 closely represents these potentially harmful high biomass dinoflagellate blooms, as discussed previously in section 3.2.5. This class displays clear seasonal variability in the St Helena Bay region (figure 4.12) with the highest percentage dominance (∼13%) occurring in summer. During spring and winter the class primarily persists in the south-western reaches of the bay, whilst the area of dominance shifts northward (to between 31 and 32°S) and slightly offshore during summer; the OWT moves closer to the coast during autumn and shows almost equal persistence (5-7%) between the northern and south-western coastal regions. This is consistent with previous studies that suggested northward movement of the coastal waters of St Helena Bay during summer when upwelling winds are strongest, whilst movement is predominantly southwards in autumn at the end of the upwelling season under prolonged periods of relaxation (Pitcher and Weeks, 2006; Fawcett et al., 2008). The years with highest bloom activity during the upwelling season (figure 4.13) were (c) 2004/2005, (g) 2008/2009 and (j) 2011/2012, consistent with previous findings using a [Chl *a*] threshold to define bloom areas (Evers-King, 2014). The 2008/2009 blooms were dominated by large dinoflagellate species *Ceratium balechii* (Pitcher and Probyn, 2011), whilst *Ceratium furca* dominated the 2011/2012 bloom (Pitcher et al., 2014); the 2004/2005 bloom will be discussed in more detail below.

Event scale classification and algorithm blending: example of a phytoplankton bloom in March 2005

A high biomass dinoflagellate bloom was recorded during a field study near Lambert's Bay in March 2005; these included toxic species such as *Dinophysis* spp. and *Prorocentrum*

reticulatum, as well as non-toxic species *Prorocentrum triestinum* (Fawcett et al., 2007). The general location of the bloom is represented by OWT6, which can be distinguished in the dominant OWT hard classification map in figure 4.14a. The membership sums (figure 4.14b) are relatively low within the regions dominated by OWT6. A possible reason for this may include problems with the atmospheric correction in hypertrophic waters. The bright pixel atmospheric correction was successfully applied over almost the entire image, which has previously been shown to be appropriate for the region and able to maintain spectral features over bloom areas (Evers-King, 2014); however, as shown earlier in section 3.3.2, there may still be spectral offsets that can lead to relatively low membership function values during the classification process. The total membership sums are resultantly much lower in the coastal region associated with the bloom. Low membership sums can also be found near the edges of clouds, whilst high memberships are associated with regions assigned to OWTs 2 to 5; due to similarities in the mean spectral shape and covariance matrices of these OWTs, as well as the unconstrained nature of the classification procedure, membership sums can be >1 in some regions.

The offshore waters are dominated by the low to medium biomass classes; figure 4.14 c and d show the weight of each of the blue-green algorithms in these regions. Much of the waters classified as OWT 5 consisted of <10 $mg.m^{-3}$, resulting in the use of the OC4E algorithm; the G2B algorithm was primarily utilized in the vicinity of the blooms, contributing between 0.1 and 0.9 of the pixel weights. The highest biomass regions apply the G3B (figure 4.14f) at weights of between 0.9 and 1. Algal2 was applied along the coast just north of the Berg River mouth and to a few small offshore filaments.

The blended Chlorophyll product (figure 4.14h) displayed values ranging from 0.02 to 329 $mg.m^{-3}$, and captures the extent of the very high biomass features in the near-shore zone off Lambert's Bay and Cape Columbine, whilst also distinguishing the various medium to low biomass filaments in the offshore regions west of 17°E. *In situ* samples taken during this time near Lambert's Bay measured [Chl *a*] between 115 and 173 $mg.m^{-3}$ (Evers-King, 2014), whilst concentrations from the blended product showed on average 118 ± 62 $mg.m^{-3}$ in this region. Within the southern Benguela the OWT classification framework and algorithm blending procedure is clearly able to provide accurate quantitative retrievals of [Chl *a*] over a very wide phytoplankton biomass range.

Switching algorithms, as used within OWTs 5 and 6, can cause discontinuities that may not be apparent just by looking at the [Chl *a*] image, but which may show up

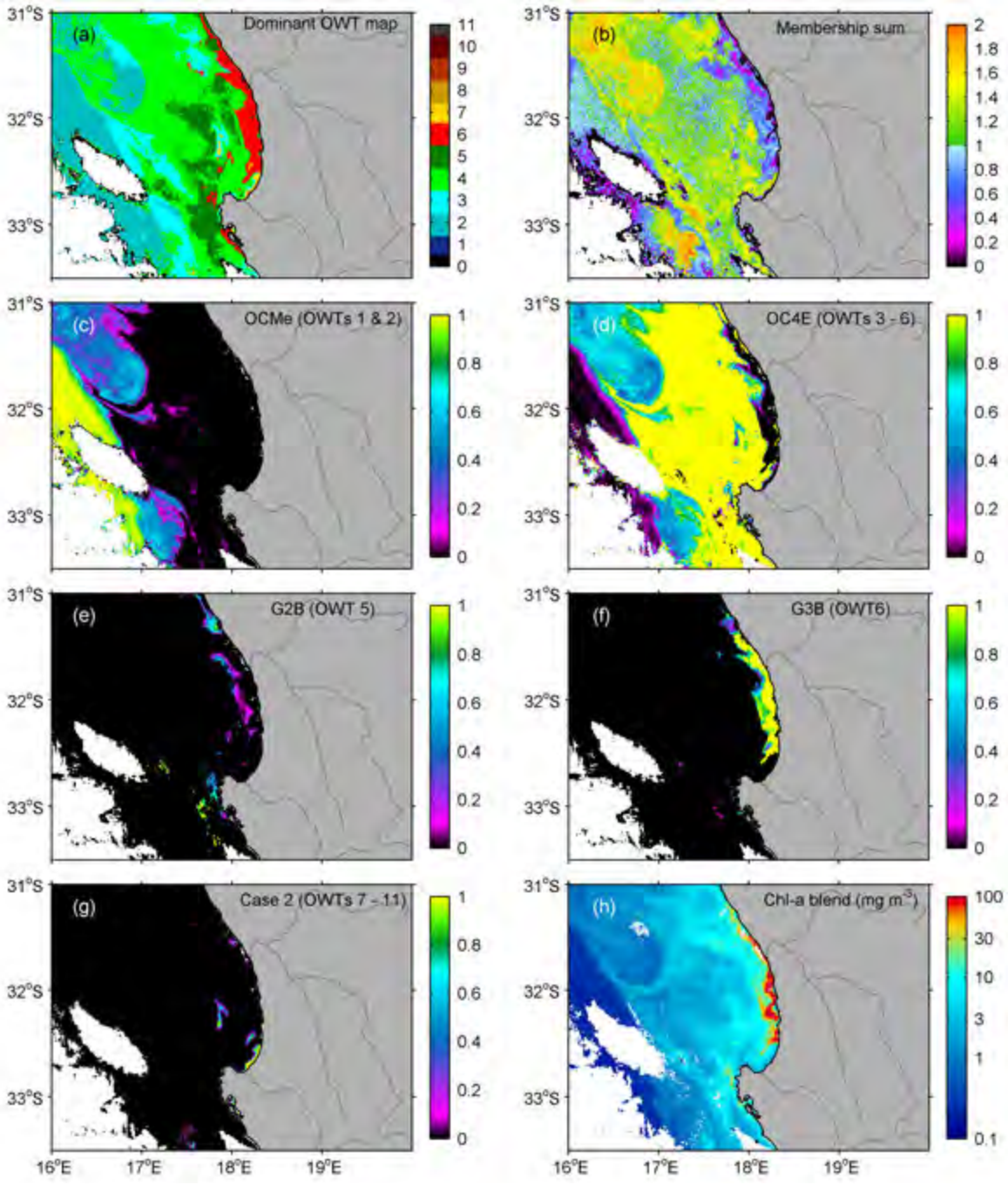


Figure 4.14: Example of the classification algorithm and chlorophyll algorithm blending procedure for 30th March 2005 in the southern Benguela. Image (a) shows the dominant (hard) OWT classification map, (b) shows the total membership sum, whilst images (c) to (g) show the weights for application of the OCMe, OC4E, G2B, G3B, and Case 2 neural network algorithms respectively. Image (h) shows the final blended [Chl *a*] product. Note that white areas represent cloud or no data.

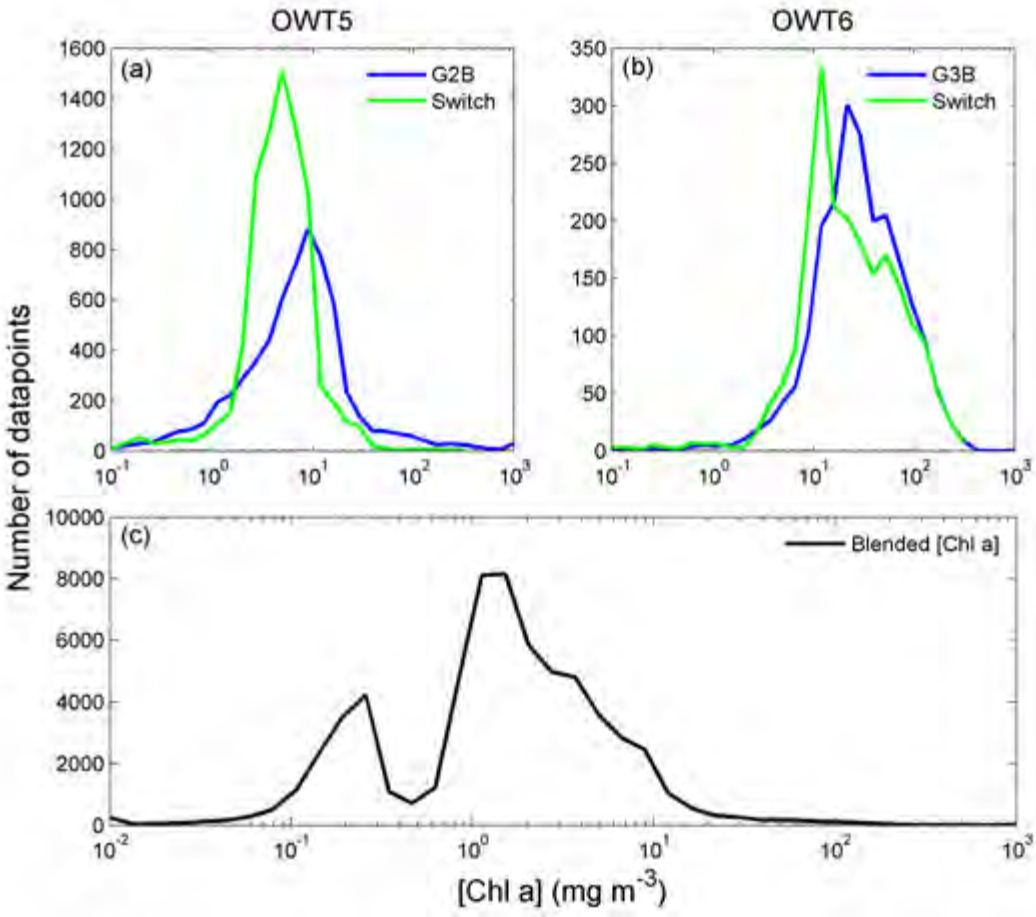


Figure 4.15: Histograms showing the [Chl *a*] of the example given in figure 4.14. Image a shows the performance of the chosen red-NIR band algorithm (G2B) and switching algorithm for all the pixels classified as OWT5, whilst image b shows the same for OWT6 (with G3B). Image c shows a histogram of the blended [Chl *a*] product across the entire image.

in histograms (e.g. Hooker et al., 1995). The appropriateness and performance of the switching algorithms utilized within OWTs 5 and 6 in figure 4.14 were assessed with histograms of the [Chl *a*] output from the appropriate red-NIR (G2B for OWT5, and G3B for OWT6) and switching algorithms; these histograms are shown in figure 4.15. Within both OWTs 5 and 6 the algorithm switch appears to produce a shift in the peak of the [Chl *a*] distribution, but does not lead to any specific or alarming discontinuity in the switched or blended products; this could suggest that these switching algorithms, although not ideal, are largely appropriate for the west coast region.

Time series analysis of the blended chlorophyll product

The performance of the blended [Chl *a*] product in the southern Benguela was assessed for three subregions as shown in figure 4.16. Spatially averaged time series of satellite [Chl *a*] products are given in figure 4.17 which included the OWT blend, Algal1 (the standard MERIS Level 2 Case 1 product), as well as two regional products: the EAP and MPH [Chl *a*]. The ranges and frequency distributions of these products in the St Helena Bay subregion are displayed in figure 4.18.

The blended product shows good coherency with the regional products during high biomass events, with additional advantages as discussed below. Due to the use of the reflectance quality flag (PCD1_13) there are slightly less data available when using the blended product compared to the EAP and MPH. The frequent application of the quality flag for the Algal1 product in the southern Benguela often results in missing data over high biomass phytoplankton blooms within the region (Evers-King, 2014); this is evident in figure 4.17, where the Algal1 product generally shows lower mean concentrations than the other regional products, particularly during the upwelling season in austral summer and autumn. Whilst the blended product has a similar shape of frequency distribution compared to Algal1 (figure 4.18), the blend shows improved ability to resolve [Chl *a*] above approximately 3 mg m^{-3}. Both the EAP and MPH are known to tend to overestimate at low concentrations (Evers-King, 2014; Matthews et al., 2012); the algorithm selection process facilitates the use of blue-green algorithms over low biomass areas, which allows the blended product to resolve [Chl *a*] below approximately 5 mg m^{-3} (figure 4.18) better than the regional products. The blended product is slightly lower for the 2005 and 2009 peaks over the St Helena Bay region (figure 4.17a), which may simply be an artefact of the selection of blue-green algorithm products over the offshore non-bloom regions; the blended product maintains very similar results to the MPH and EAP for the two smaller coastal subregions. Resultantly the algorithm blending procedure provides extended range of applicability in the southern Benguela compared to standard and regional satellite products.

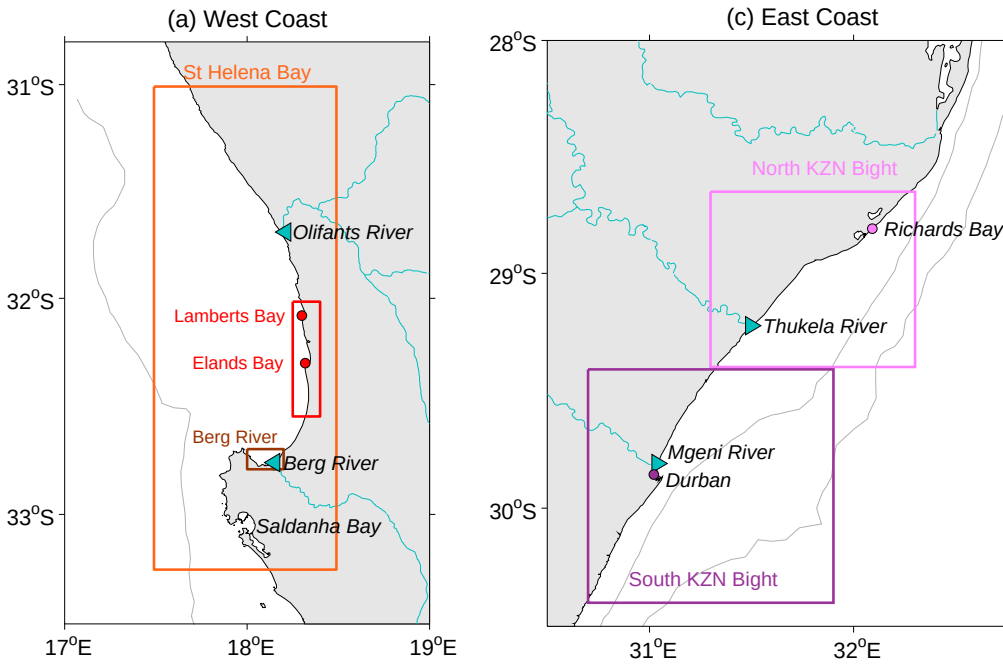


Figure 4.16: Schematic map outlining the bounding boxes used for creating spatial averages in St Helena Bay (orange) as well as the Elands/Lambert's Bay (red) and Berg River subregions (brown).

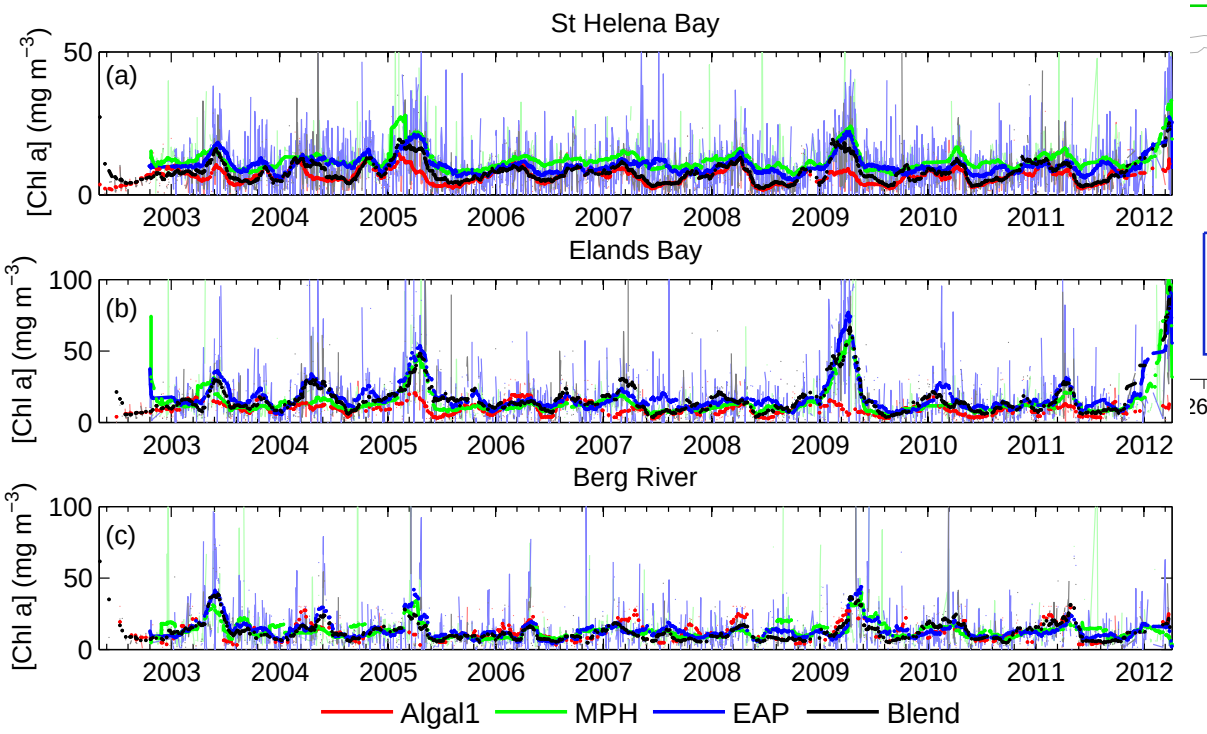


Figure 4.17: Time series of [Chl a] (mg m^{-3}) for the (a) St Helena Bay region, as well as the (b) Elands/Lambert's Bay and (c) Berg River subregions as detailed in figure 4.16. The products shown include Algal1 (red), MPH (green), EAP (blue) and the blended product (black). A 30 day running mean is shown by the bold lines, whilst the full time series is shown by the lines with lighter shades.

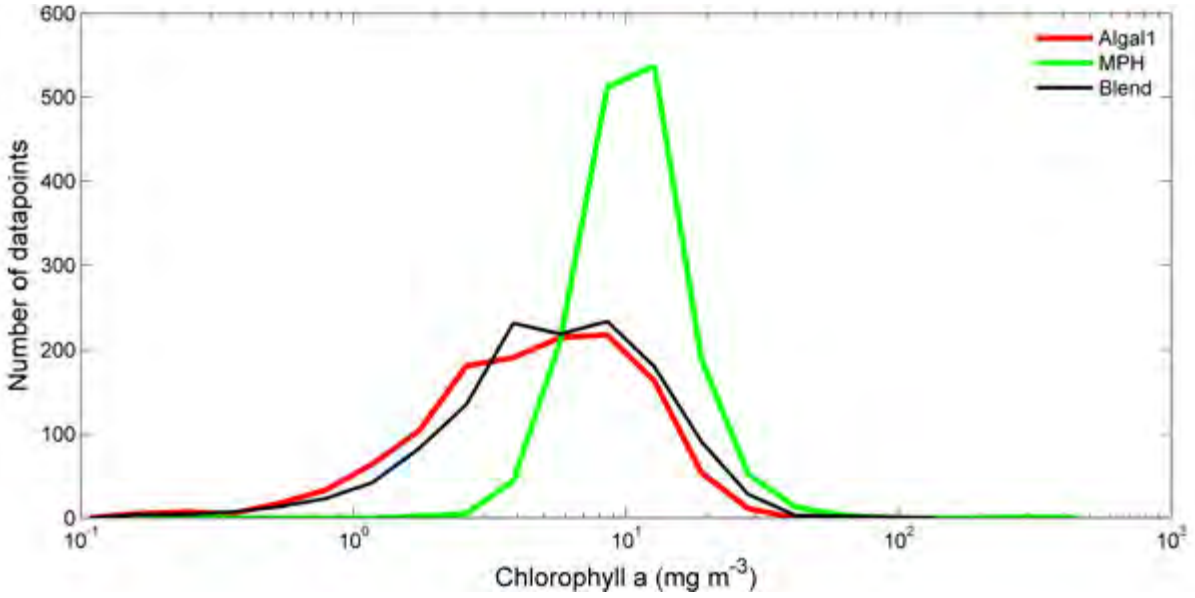


Figure 4.18: Histograms of the time series of daily [Chl *a*] products (as shown in figure 4.17) for the St Helena Bay region (as detailed in figure 4.16). The products shown include Algal1 (red), MPH (green) and the blended product (black).

4.3.4 Case study 2: Algoa Bay and Agulhas Bank region

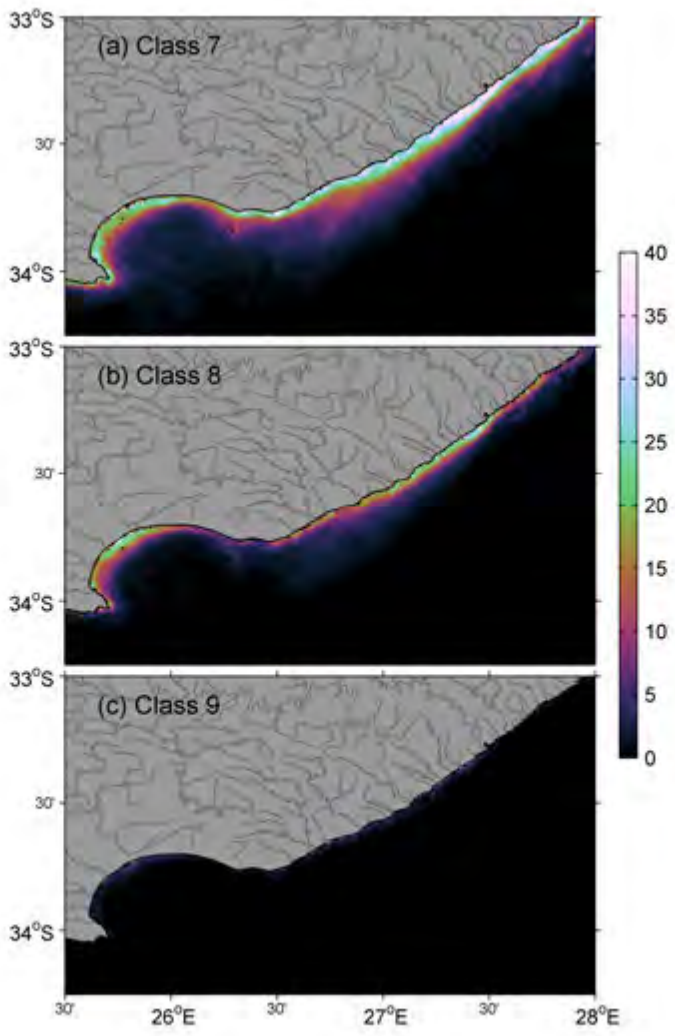


Figure 4.19: The percentage dominance of (a) OWT7, (b) OWT8, and (c) OWT9 in Algoa Bay for MERIS data from 2002 through 2012. The dark grey lines represent all the rivers in the region.

High wave energy and fine sand along the south coast produce wide gently sloping dissipative beaches with offshore bars (Cooper, 2001) which can result in re-suspension of bottom sediments or even the effects of bottom reflectance, potentially contributing to the presence of the elevated reflectance OWTs in this region (figure 4.19). Both OWTs 7 and 8 have the farthest offshore extent just upstream of Algoa Bay, a region where meanders in the Agulhas current may entrain and transport particles offshore (Jackson et al., 2012). OWTs 7 to 9 show some spatial coherence to the location of estuaries along the coast that contribute fluvial sediments to the surface waters; most of the fine sediments from rivers in this region are carried off in suspension by the Agulhas Current

(Flemming, 1981). OWT7 shows the highest persistence on the coastline northwest of Algoa Bay where several estuaries are found, whilst both OWTs 8 and 9 show elevated persistence near the Sunday River on the north-western shores of Algoa Bay as well as downstream of the Keiskamma River (27.5°E)

Along the south coast the highest surface [Chl *a*] are typically found near the coast, particularly between Cape Agulhas and Algoa Bay (Shannon et al., 1984; Demarcq et al., 2003); these regions were consistently classified as the moderate to high biomass OWT5. Increased persistence of OWT5 during spring and autumn (figure 4.20) is consistent with the initiation of the spring bloom on the Agulhas Bank (Mitchell-Innes et al., 1999) and the peak of the coastal upwelling season respectively; this bimodal trend has been previously observed in satellite time series (Jarre et al., 2015). Slightly lower persistence in OWT5 is seen in summer when thermal stratification prompts the formation of subsurface [Chl *a*] maxima over much of the Agulhas Bank in close proximity to the thermocline (Probyn et al., 1994) when nutrients supply to the surface waters is limited (Shannon et al., 1984). There is a slight 'bulge' in the offshore persistence of the high biomass OWT5 in spring and autumn off Algoa Bay, possibly associated with meanders in the Agulhas Current and associated upwelling (Goschen and Schumann, 1988; Lutjeharms et al., 2000a) that occur at this longitude. Similar to the results for the west coast the persistence of OWT4 appears to be linked to areas associated with dynamic upwelling at the inner edge of the Agulhas Current (Lutjeharms et al., 2000a) between 26 and 28°E and in the lee of the wind-driven coastal upwelling zones at the prominent Capes (Schumann et al., 1982). The increased persistence of moderate biomass OWT4 to the east of Algoa Bay during autumn and winter (figure 4.21) could potentially be attributed to a more turbulent environment from deeper water column mixing during these months (Beckley, 1988).

Studies have shown that the community structure in the coastal waters of the Agulhas Bank are often dominated by diatoms (Barlow et al., 2010; Van der Lingen et al., 2016), which favour the more dynamic and nutrient-rich environment that upwelling offers. Although the region has a much lower occurrence of HABs than the west coast, and shows a persistence of <2% of OWT6 (not shown), dinoflagellates such as *Gymnodinium* cf. *mikimotoi*, *Prorocentrum rostratum*, *Prorocentrum triestinum* and *Scrippsiella trochoidea* have occasionally occurred in the region (Pitcher and Calder, 2000) and frequent blooms of *Noctiluca miliaris* are common during summer and autumn whilst *Lingulodinium polyedrum* has been prevalent along the south coast during the summer months since 2014 (Tommy Bornman, personal communication, February 25, 2016). An

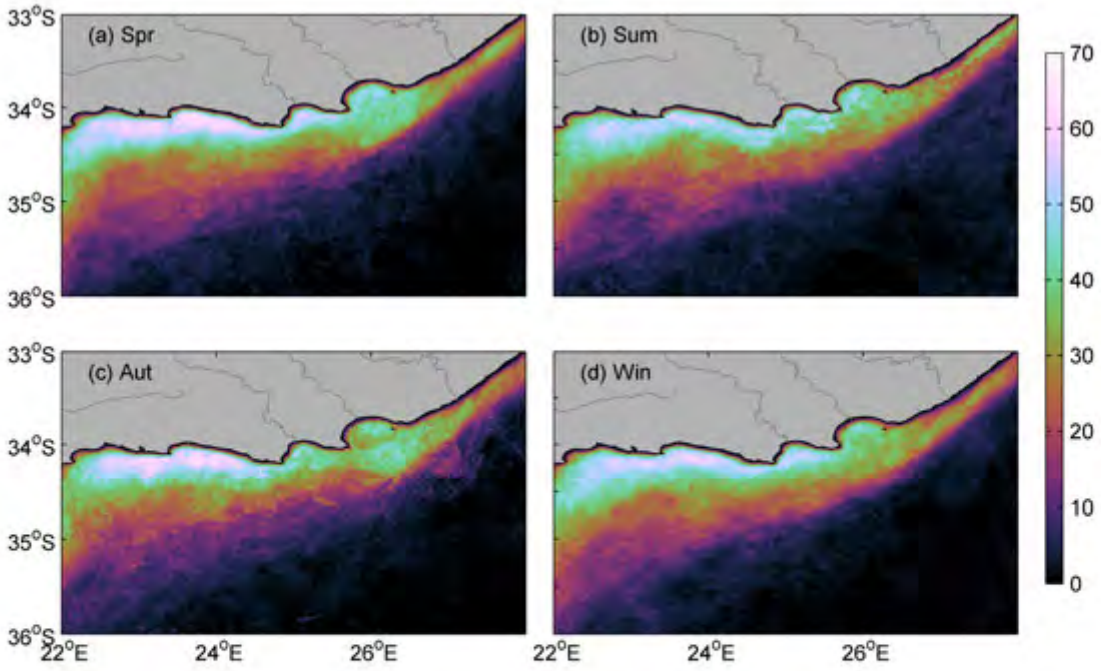


Figure 4.20: The percentage seasonal dominance of OWT5 over the eastern Agulhas Bank for MERIS data from 2002 through 2012.

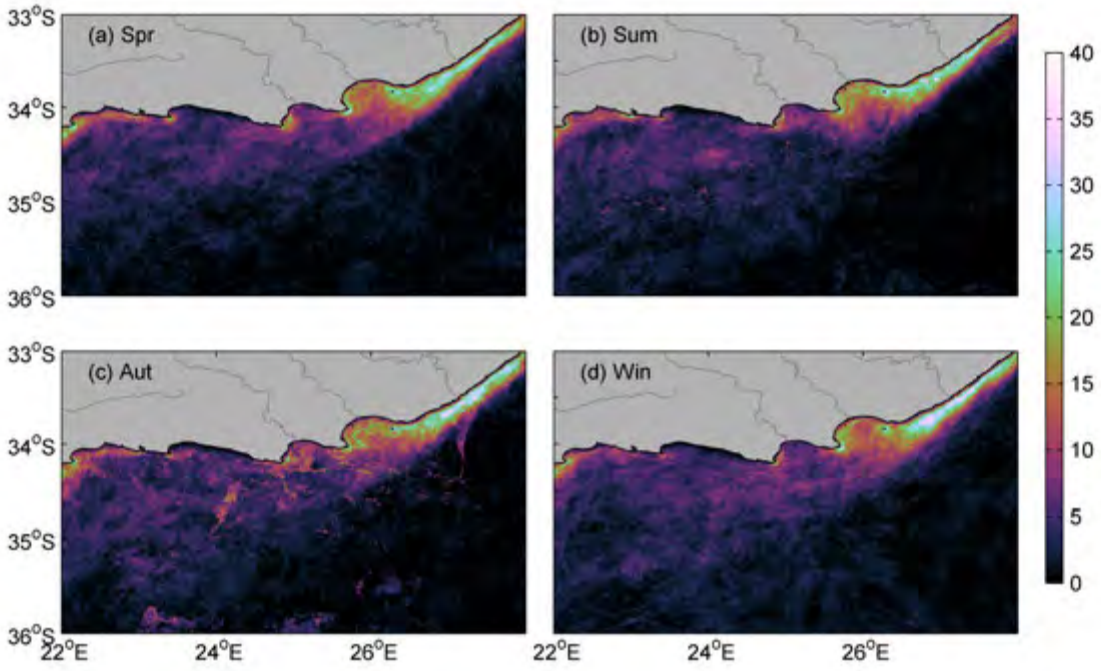


Figure 4.21: The percentage seasonal dominance of OWT4 over the eastern Agulhas Bank for MERIS data from 2002 through 2012.

example of a high biomass bloom event of unknown assemblage is given below.

Event scale classification and algorithm blending: example of a phytoplankton bloom in April 2004

High biomass conditions were observed over the south coast at the coast as well as over the offshore Agulhas Bank region during April 2004. Similar high biomass conditions were observed in MODIS Aqua imagery on this day (not shown), whilst the regional extent and persistence of this bloom was also seen in Jarre et al. (2015). Figure 4.22 shows an example of the conditions for the 6th of April 2004.

The dominant OWT map (figure 4.22a) shows the presence of the high biomass OWT 5 stretching from Algoa bay westward and offshore on the continental shelf region along the inner edge of the Agulhas Current, with small patches of the very high biomass OWT6 near the coast. Low biomass OWT 2 dominates over the offshore Agulhas Current region, whilst a circulation feature incorporating six different OWTs can be distinguished along the shelf and current edges. Although the Agulhas follows a stable trajectory along the east coast, it has a tendency to meander as the continental shelf widens (Lutjeharms et al., 2000a) downstream of approximately 27°E; these meanders often form cyclonic eddies (Lutjeharms et al., 2000a) which can entrain and transport particles offshore (Jackson et al., 2012).

Membership sums (figure 4.22b) are lowest near the coast and over the high biomass areas, most likely due to atmospheric correction problems over these waters. The BPAC was successfully applied over almost the entire image and all bloom areas; low memberships over bloom areas could potentially be due to the spectral offset resulting from BPAC as seen previously along the west coast blooms in section 4.3.3. Atmospheric correction failure is also evident from the white patches (figure 4.22h) to the west of the bloom near the coast where the PCD1_13 flag has been raised; upon closer inspection of the flagged pixels it was found that they were raised due to uncertain aerosol type and optical thickness.

The OCMe and OC4E are applied over most of the offshore oceanic regions, whilst much of the shelf region and the circulation feature uses the OC4E algorithm (figure 4.22d). The G2B algorithm has high weights over the bloom region and blends with the G3B at the locations of the highest biomass. Thin filaments of Case 2 waters (figure 4.22g) can be seen within the circulation feature, possibly from entrainment of sediment-containing

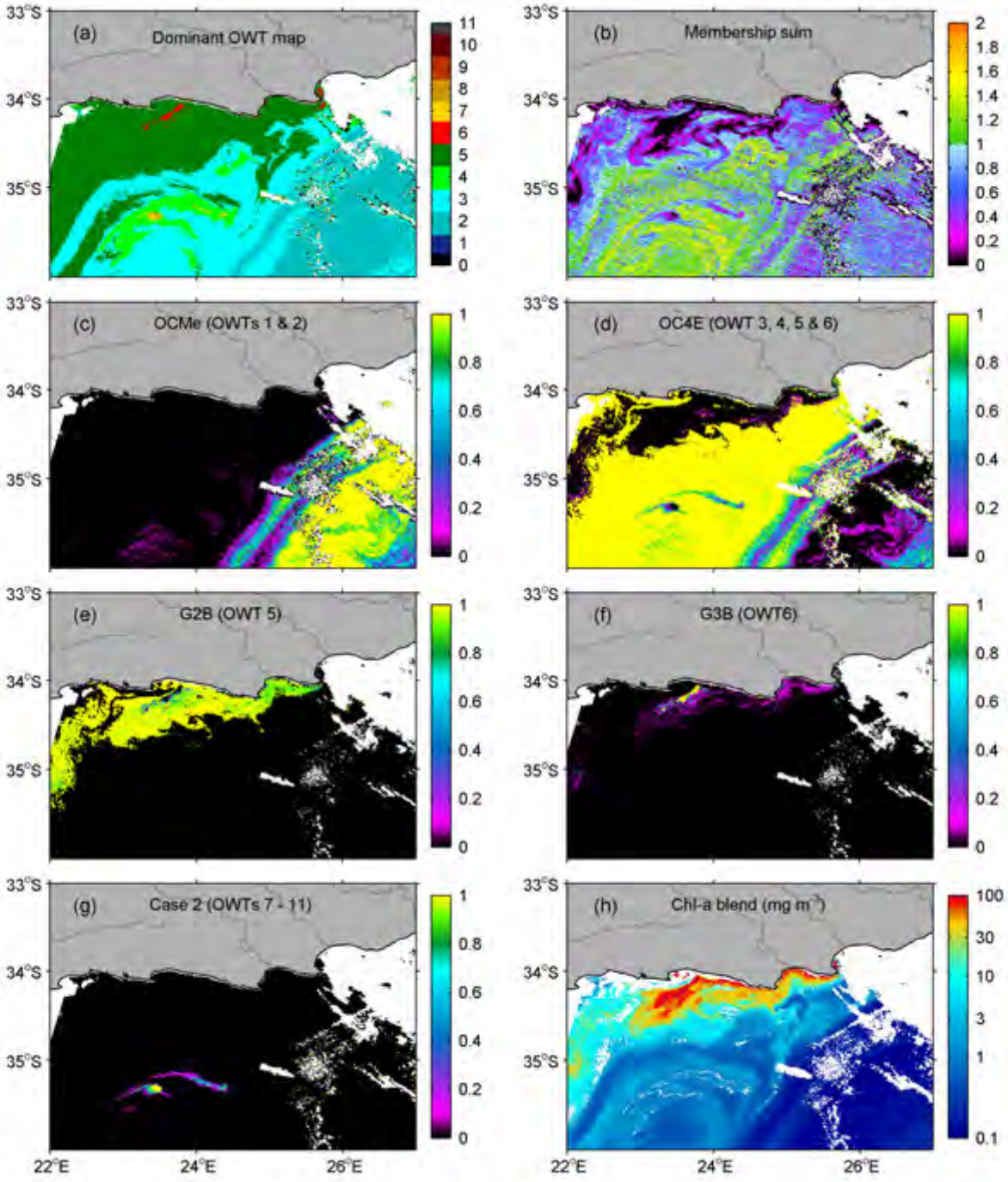


Figure 4.22: Example of the classification algorithm and chlorophyll algorithm blending procedure for 6th April 2004 along the south coast. Image (a) shows the dominant (hard) OWT classification map, (b) shows the total membership sum, whilst images (c) to (g) show the weights for application of the OCMe, OC4E, G2B, G3B, and Case 2 neural network algorithms respectively. Image (h) shows the final blended [Chl *a*] product. Note that white areas represent cloud or no data.

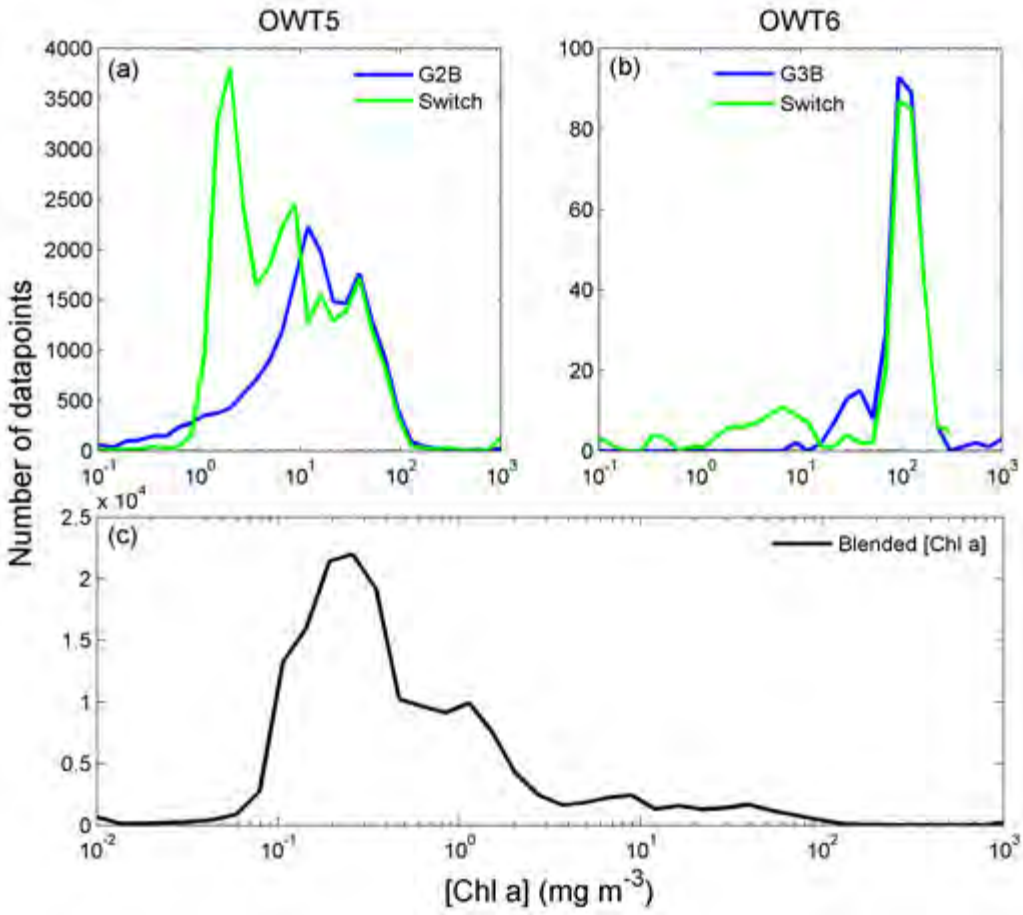


Figure 4.23: Histograms showing the [Chl *a*] of the example given in figure 4.22. Image a shows the performance of the chosen red-NIR band algorithm (G2B) and switching algorithm for all the pixels classified as OWT5, whilst image b shows the same for OWT6 (with G3B). Image c shows a histogram of the blended [Chl *a*] product across the entire image.

coastal water, although the low membership sum over the filament may indicate difficulties with the atmospheric correction with subsequent potentially erroneous classification. The bloom can be seen in the blended chlorophyll product (figure 4.22h), and appears to be located primarily in the coastal zone between Algoa Bay and approximately 24°E, after which it gets advected farther offshore over the shelf region. Due to the switching that occurs at 10 mg m^{-3} within OWT 5, there is often a hard boundary that occurs at the edge of the phytoplankton bloom; however this does not appear to negatively impact the quality of the blended chlorophyll product with regard to synoptic coherency.

As in the previous section, the appropriateness and performance of the switching algorithms utilized within OWTs 5 and 6 in figure 4.22 were assessed with histograms of the [Chl *a*] output from the appropriate red-NIR (G2B for OWT5, and G3B for OWT6) and

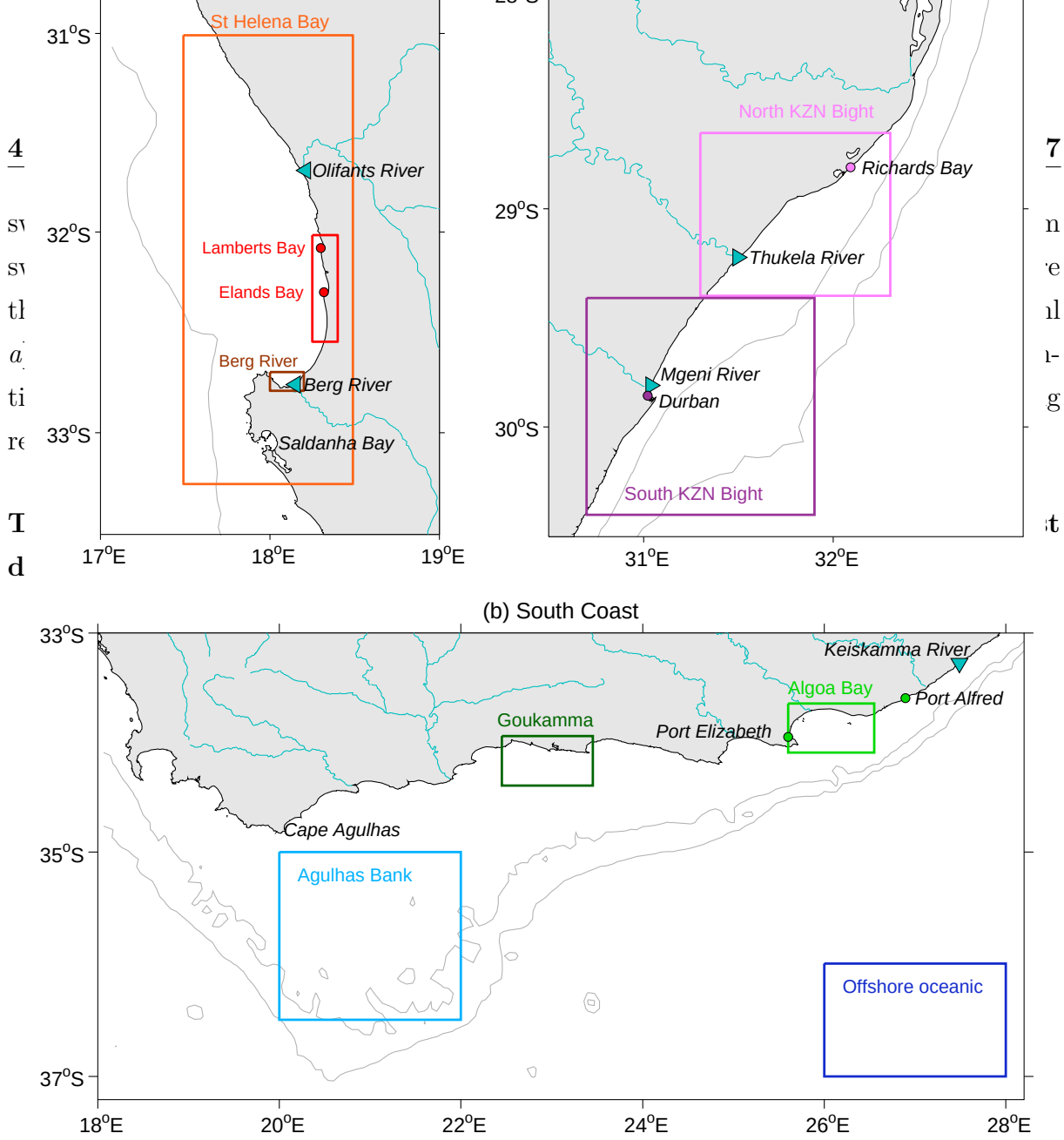


Figure 4.24: Schematic map outlining the bounding boxes used for creating spatial averages along the south coast, with Algoa Bay (light green), Goukamma coast (dark green), Agulhas Bank (light blue), and offshore oceanic (dark blue) subregions

Four subregions (figure 4.24) were designated for time series analysis within the south coast domain in order to assess the performance of the blended product in subregions that experience different environmental conditions and physical forcing mechanism. These sites included a zone of intermittent coastal wind-driven upwelling (Goukamma coast, shown in dark green), a bay near a dynamic upwelling zone (Algoa Bay, shown in light green), continental shelf zone (Agulhas Bank, shown in light blue), and an offshore oceanic region (shown in dark blue). A spatially averaged time series of satellite [Chl *a*] products for the south coast subregions are shown in figure 4.25, with the corresponding frequency distributions for all these regions combined in figure 4.26. Since no regional products are available for this area, the [Chl *a*] products include only the OWT blend, Algal1 and

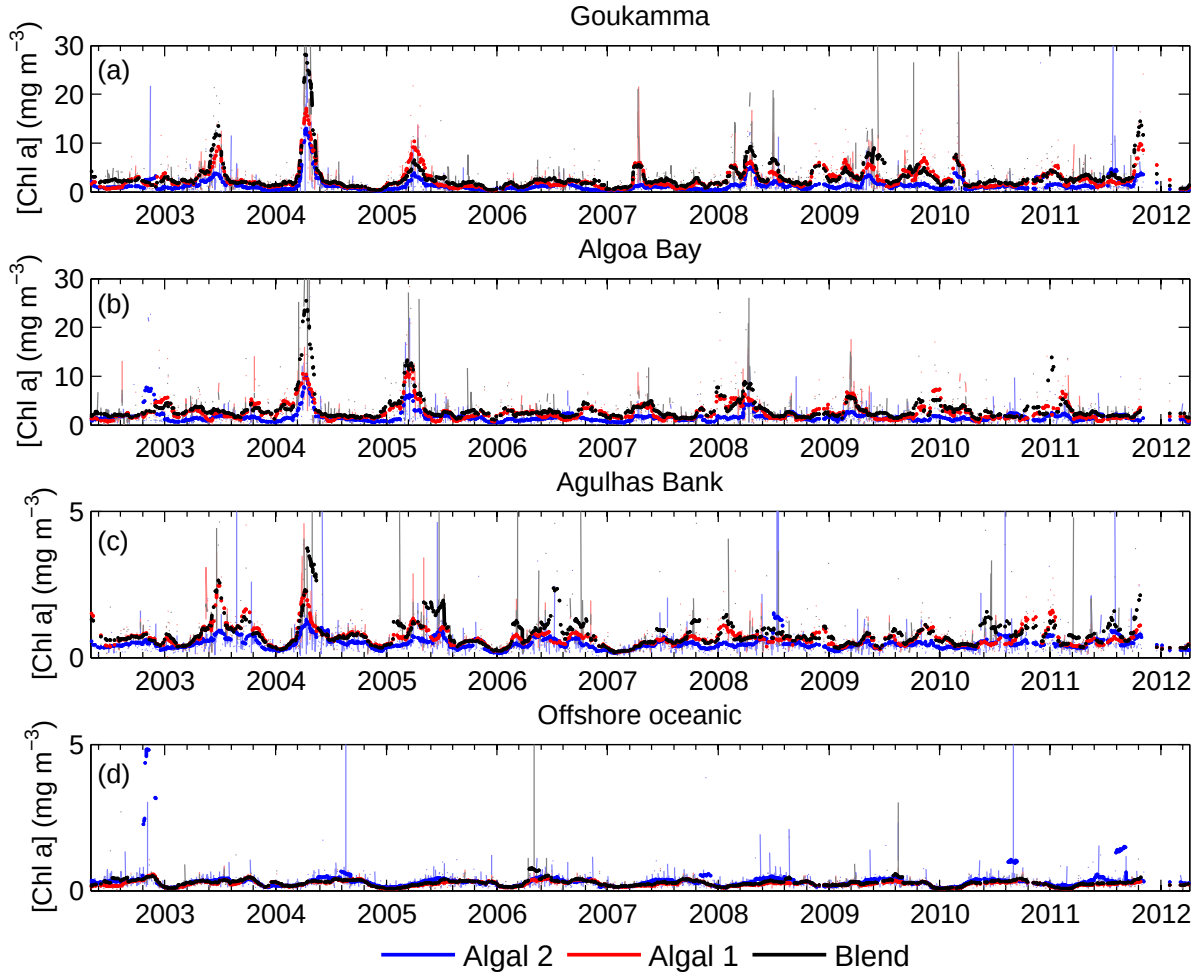


Figure 4.25: Time series of [Chl *a*] (mg m^{-3}) for the (a) Goukamma, (b) Algoa Bay, (c) Agulhas Bank and (d) offshore oceanic regions as detailed in figure 4.24. The products shown include Algal2 (blue), Algal1 (red) and the blended product (black). A 30 day running mean is shown by the bold lines, whilst the full time series is shown by the lines with lighter shades.

Algal2 (the standard MERIS Level 2 Case 2 product).

The blended product showed good coherency with the two MERIS products in all subregions, with additional advantages as shown below. Both the Goukamma (figure 4.25a) and Algoa Bay (figure 4.25b) subregions are affected, to some degree, by nutrient enrichment through upwelling, which may result in the surface expression of increased levels of phytoplankton biomass in these regions; these peaks showed synoptic coherence with MODIS data over this period (Jarre et al., 2015). These nutrient-rich waters from coastal upwelling may occasionally be advected into adjacent bays or offshore onto the Agulhas Bank (Walker, 1986; Lutjeharms, 2006) and could lead to higher biological productivity on the shelf (Lutjeharms et al., 2000a; Lutjeharms, 2007), resulting in the peaks in average

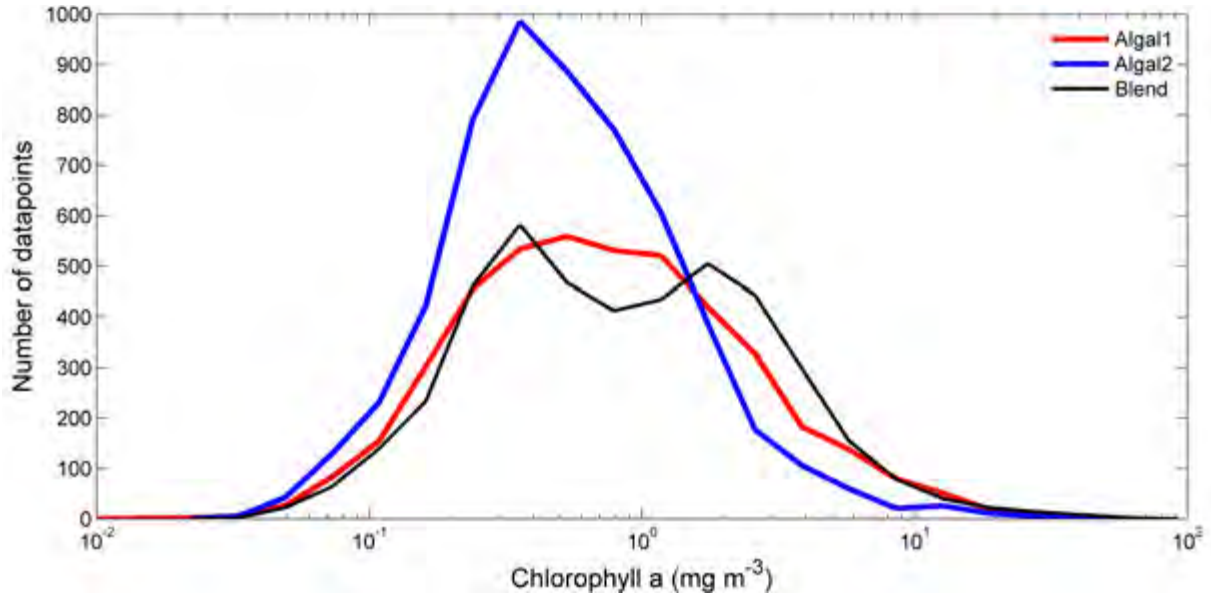


Figure 4.26: Histograms of the time series of daily [Chl *a*] products (as shown in figure 4.25) for all the south coast subregions (as detailed in figure 4.24). The products shown include Algal1 (red), Algal2 (blue) and the blended product (black).

[Chl *a*] as seen during 2003 and 2004 (figure 4.25c). The blended product demonstrates the ability to capture the large range of [Chl *a*] that may be associated with upwelling-induced phytoplankton bloom events in the coastal and shelf environments of the south coast, whilst also being able to deal with the small range of [Chl *a*] that may be expected offshore over the open ocean (figure 4.25d). The blended product demonstrates a broader range of application and improved capability to resolve [Chl *a*] above approximately 2 mg m^{-3} compared to Algal1 and Algal2 (figure 4.26). It is possible that the bi-modal nature of the blended [Chl *a*] product may be an artefact of the switching algorithm used within OWT5.

4.3.5 Case study 3: KwaZulu-Natal Bight, east coast

The east coast of South Africa has a relatively featureless coastline which enables the warm, fast-moving Agulhas Current to flow in close proximity to the coast. A major coastal feature is the KZN Bight which has a broader shelf environment compared to the adjacent coastline. The Agulhas Current moves farther offshore in this region producing lee eddies along its inshore edge (Lutjeharms, 2006) that create a more retentive shelf environment compared to the adjacent coastal regions.

Due to the subtropical climate of the region, river runoff and terrestrial nutrient input into the nearshore zone may be increased during summer (Barlow et al., 2015). This seasonal signal is represented in the elevated reflectance of OWTs 7 and 8, presented in figure 4.27, which show the highest percentage persistence in the coastal regions north of the Thukela River (indicated on the map in figure 4.30) during summer when there is higher fluvial discharge into the bight waters. Flemming (1981) found that there is a general lack of large mud deposits on the shelf of the east coast, which would imply that most of the fine materials like silts and clays from the rivers are carried off in suspension by the currents. Alongshore currents north of the Thukela may provide a mechanism for the movement of turbid waters and suspended sediments northwards along the coast. The wide offshore extent and the high persistence of these elevated reflectance water types highlights the optical complexity and frequent need for Case 2 algorithms to resolve [Chl *a*] in these shelf waters.

The mesotrophic nature of the KZN Bight (Barlow et al., 2008) is confirmed by the persistence of the moderate biomass OWT4 over the shelf. The presence of this OWT appears to be closely related to nutrient availability and upwelling since it shows the highest persistence over the northern Bight where a persistent upwelling cell in the vicinity of St Lucia provides a source of nutrients that are subsequently transported southward over the KZN Bight shelf (Meyer et al., 2002). A sharp decrease in persistence at the shelf edge delineates the strong oceanic influence of the oligotrophic Agulhas Current. There appears to be some seasonal variability in OWT4 with the highest persistence (up to 45%) during summer (figure 4.27d), which may potentially be linked to increased nutrient input from riverine influxes. OWT4 also shows persistence off Durban in the south throughout the year, where elevated phytoplankton biomass has been recorded in areas of cooler water and increased nutrient availability (Barlow et al., 2008, 2010) which has been linked to the persistent lee eddy off Durban (Lamont and Barlow, 2015).

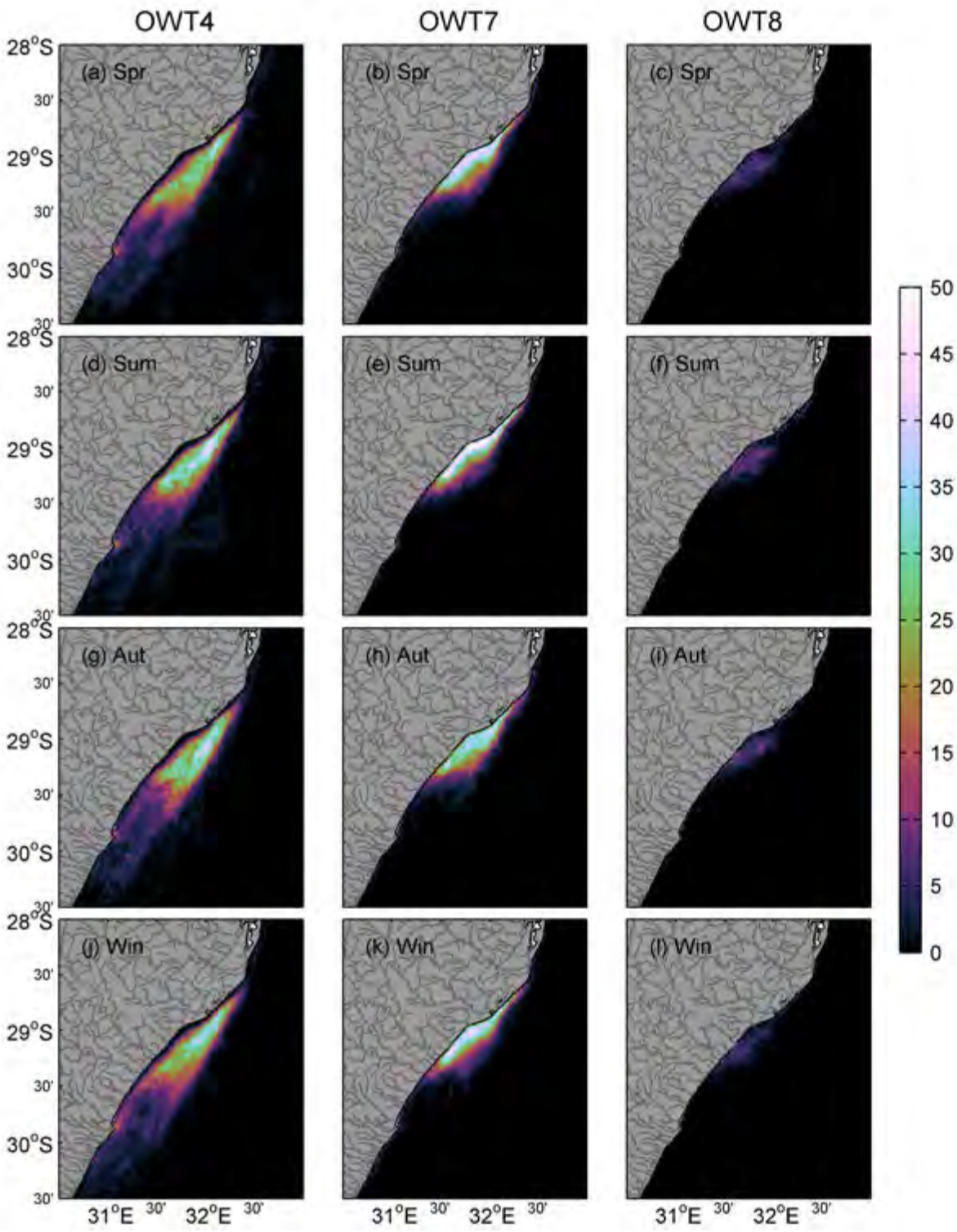


Figure 4.27: The percentage seasonal dominance of OWT4 (panels a, d, g, and j), OWT7 (panels b, e, h, and k) and OWT8 (panels c, f, i, and l) in the KwaZulu-Natal Bight for MERIS data from 2002 through 2012. The dark grey lines represent all the rivers in the region.

Event scale classification and algorithm blending: example of high Thukela flow in March 2004

Relatively high flow rates were observed for the Thukela River (located in the middle of the KwaZulu-Natal Bight at approximately 29.22°S 31.50°E) during the summer of 2004; during the two weeks leading up to the 9th of March (the date of the example image shown in figure 4.28) the average daily flow rate was approximately 184 ± 44 m^3s^{-1}, whilst median daily flow rate for the Thukela River over the course of the MERIS mission was approximately 27 m^3s^{-1}. Large flood events could have a substantial influence on the shelf waters (Meyer et al., 2002) from the influx of terrigenous particulate matter and dissolved organic substances, whilst also leading to localized increases of nutrients and phytoplankton (Carter and Schleyer, 1988; Barlow et al., 2015).

The low biomass OWTs 1 and 2 dominated in the offshore zone and southern KZN Bight (figure 4.28a), whilst the elevated reflectance classes dominated at the Thukela River mouth, as well as in the coastal zone north of the mouth. The bi-directional alongshore currents in the coastal waters north of the Thukela River (Schumann, 1987) provide a mechanism for the movement and retention of sediment-laden waters in the near-coastal zone. The low to moderate biomass OWTs 3 and 4 were located on the offshore edge of the elevated reflectance classes; this could represent the zone where some of the particulate matter had already sunk out of the water column, whilst the nutrients were still available to phytoplankton. A small strip of OWT5 was also visible in the southern half of the bight running parallel to the coast. The membership sums (figure 4.28b) were relatively low over the region, with the lowest sums found at the cloud edges and along the coast; the coastal zones south of the Thukela River and to the north of the KZN Bight show a larger offshore extent of low membership sums than the northern half of the bight. Regions of OWTs 8 to 11 also have very low membership. Highest memberships occur in the offshore oceanic zone where pixels potentially have mixed membership to low biomass OWTs.

The OCMe algorithm (figure 4.28c) was applied over most of the oceanic areas. Large solitary meanders known as Natal pulses occasionally form in the KZN Bight consisting of cyclonic eddies on the landward edge of the Agulhas Current (Lutjeharms and Roberts, 1988); an example of such a circulation feature can be seen in figure 4.28 c and d where the two blue-green algorithms are blended together. The two red-NIR algorithms had very limited application in the example, with G2B (figure 4.28e) limited to the southern KZN Bight whilst OWT6 (G3B) was absent. Solely Algal2 (figure 4.28g) is applied over much of the northern KZN Bight and most of the surrounding near-shore zone; OC4E is

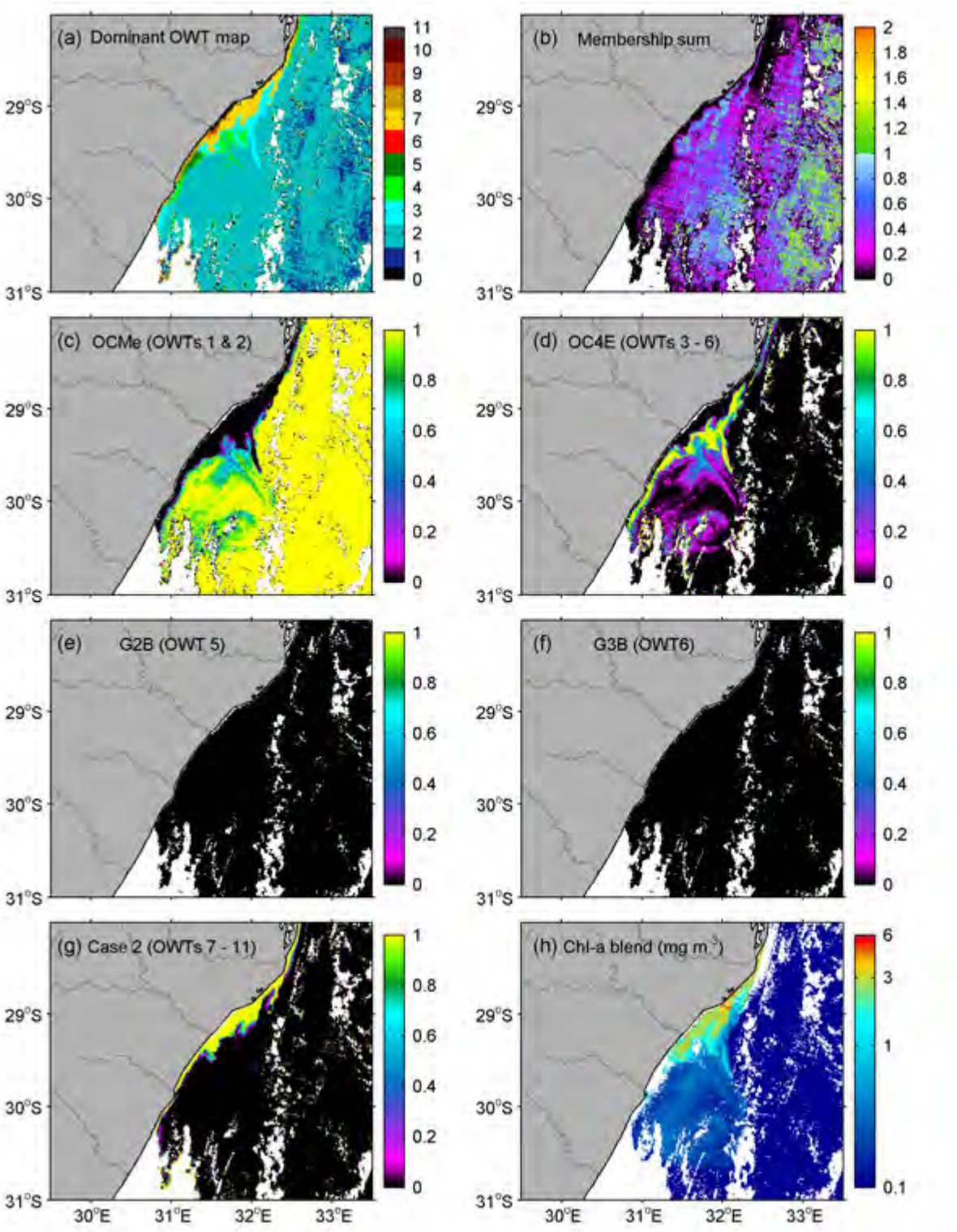


Figure 4.28: Example of the classification algorithm and chlorophyll algorithm blending procedure for 9th March 2004 in the KwaZulu-Natal Bight. Image (a) shows the dominant (hard) OWT classification map, (b) shows the total membership sum, whilst images (c) to (g) show the weights for application of the OCMe, OC4E, G2B, G3B, and Case 2 neural network algorithms respectively. Image (h) shows the final blended [Chl *a*] product. Note that white areas represent cloud or no data.

applied in the small offshore filaments and along the edges of the OWT7 zone.

The blended [Chl *a*] is shown in figure 4.28h, where higher [Chl *a*] can be seen in the coastal regions of the northern KZN Bight. [Chl *a*] appear to be slightly lower in the areas dominated by OWTs 8 and over which are most likely to represent zones with high concentrations of fluvial particulate matter; studies have suggested that these conditions may inhibit phytoplankton productivity on the bight through light limitation (Carter and d'Aubrey, 1988; Carter and Schleyer, 1988). The relatively high concentrations of phytoplankton in the coastal regions appear to be advected offshore by the cyclonic circulation feature toward the south-east of the bight, and incorporated into the gyre; Jackson et al. (2012) proposed this mechanism for entrainment and offshore movement of particulate matter through a Natal Pulse for the Agulhas Bank region.

In figure 4.28h, many of the coastal regions showing low membership sums have been flagged by PCD1_13 and resultantly masked out; these represent regions that likely experienced problems with the atmospheric correction. Within this image the bright pixel atmospheric correction was only successfully applied over the northern KZN Bight regions dominated by elevated reflectance classes; this means that in all other regions the residual marine signal in the NIR was set to that of pure seawater (Lerebourg and Bruniquel, 2011), which could lead to an underestimation of the water-leaving reflectance in the shorter wavelengths (Smith et al., 2013). Most of the areas dominated by OWT5 were flagged by PCD1_13; since this class has the overall lowest spectral reflectance, it is possible that an overestimation of the aerosol correction (leading to erroneously low or even negative reflectance in the blue spectral regions) could potentially produce satellite reflectance spectra that would have the closest membership (albeit small) to OWT5.

As in the previous section, the appropriateness and performance of the switching algorithms utilized within OWTs 5 and 6 in figure 4.28 were assessed with histograms of the [Chl *a*] output from the appropriate red-NIR (G2B for OWT5, and G3B for OWT6) and switching algorithms; these histograms are shown in figure 4.29. Very few pixels were assigned to either OWTs 5 or 6, and there were no apparent discontinuities introduced by the algorithm switch. The high persistence of the reflectance quality flag over the east coast region (as seen previously in figure 4.5) may result in pixels being falsely assigned to OWT5, leading to erroneously high [Chl *a*] from the G2B algorithm as shown in figure 4.29a; thus the switch is largely appropriate for this region from a quality control perspective.

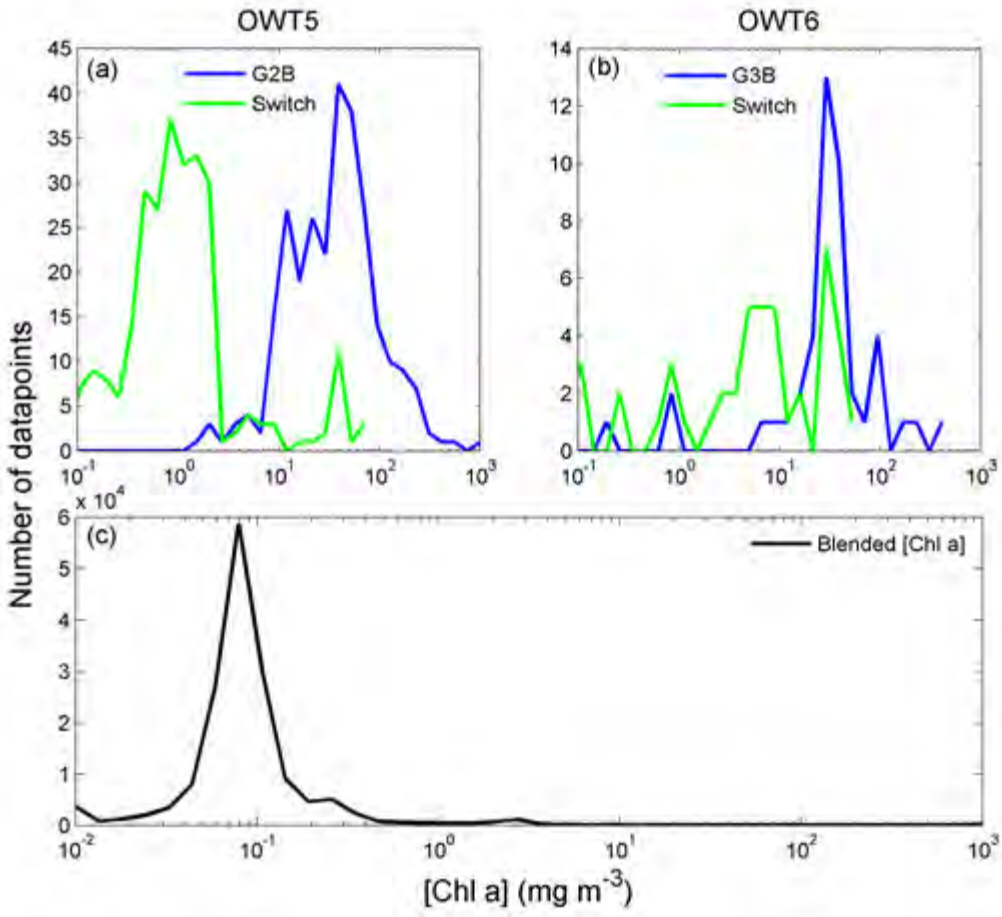


Figure 4.29: Histograms showing the [Chl *a*] of the example given in figure 4.28. Image a shows the performance of the chosen red-NIR band algorithm (G2B) and switching algorithm for all the pixels classified as OWT5, whilst image b shows the same for OWT6 (with G3B). Image c shows a histogram of the blended [Chl *a*] product across the entire image.

Time series analysis of the blended chlorophyll product within the east coast domain

Two subregions (figure 4.30) were designated for time series analysis within the east coast domain including the north and south of the KZN Bight. Spatially averaged time series of satellite [Chl *a*] products for these subregions are shown in figure 4.31, with the corresponding frequency distributions for all both regions combined in figure 4.32. Since no regional products are available for this area, the [Chl *a*] products include only the OWT blend, Algal1 and Algal2 (the standard MERIS Level 2 Case 2 product).

The satellite [Chl *a*] are consistent with *in situ* measurements from Carter and Schleyer (1988) who reported 0.03-3.9 mg m^{-3} in the KZN Bight. The blended product shows more similarity to the Algal2 product in the northern KZN Bight site, whilst the Algal1

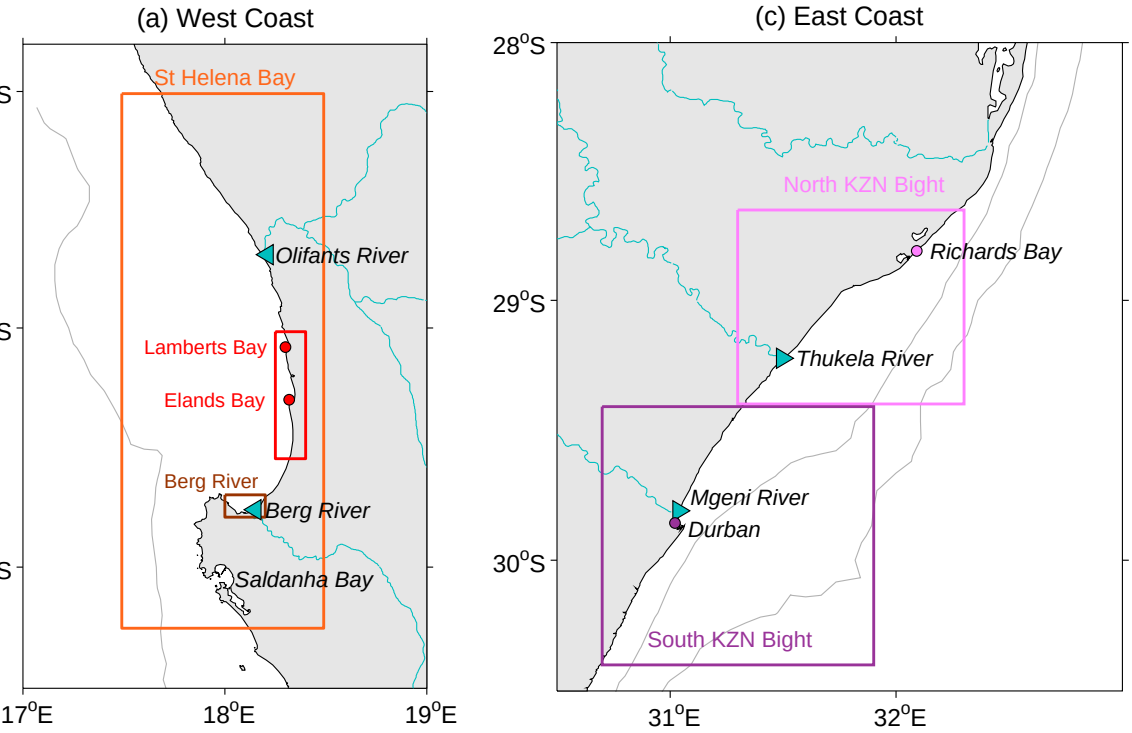


Figure 4.30: Schematic map outlining the bounding boxes used for creating spatial averages for the KwaZulu-Natal (KZN) bight, showing the northern(pink) and south(purple) KZN Bight subregions.

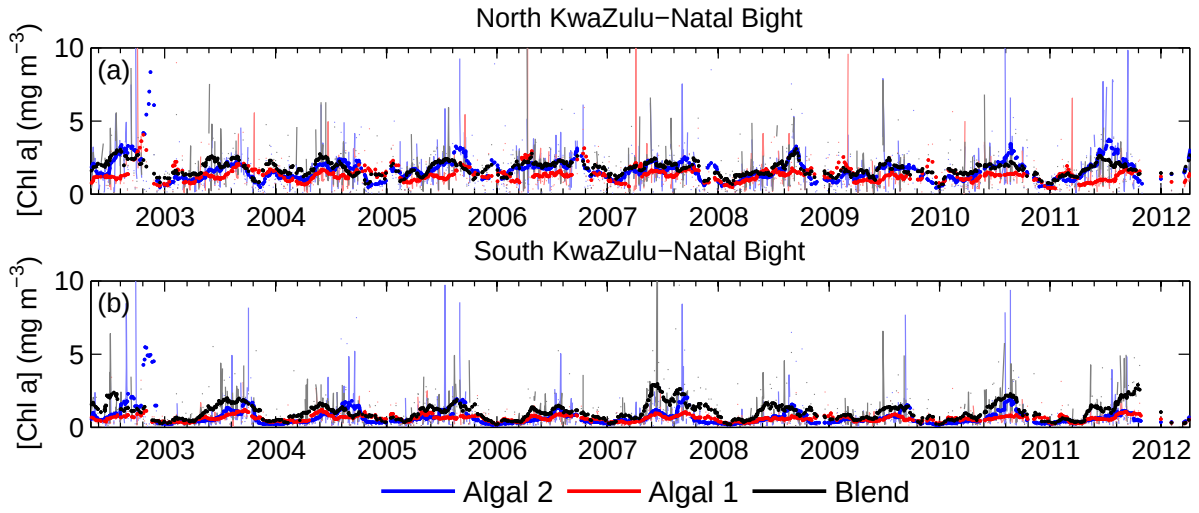


Figure 4.31: Time series of [Chl *a*] (mg m^{-3}) for the (a) northern KwaZulu-Natal and (b) southern KwaZulu-Natal Bight regions as detailed in figure 4.30. The products shown include Algal2 (blue), Algal1 (red) and the blended product (black). A 30 day running mean is shown by the bold lines, whilst the full time series is shown by the lines with lighter shades.

product generally shows lower mean [Chl *a*]. The high dominance of elevated reflectance OWTs 7 and 8 may result in the selection of Algal2 for the blended product, as well as the exclusion of data from the Algal1 product due to quality flagging which can lead to lower average concentrations as previously seen in the west coast subregions. Although the blended product shows good coherency with the MERIS products with regard to the seasonal signal over the southern bight, the blended product appears to show generally higher [Chl *a*] than Algal1 and Algal2, particularly during austral winter. This may be an artefact of the persistence of OWT5 over the southern bight. However, as demonstrated in figure 4.28a, spectra may occasionally be erroneously classified as OWT5 in the east coast region as a result of difficulties with the atmospheric correction; this could subsequently increase the risk of inappropriate application of the G2B algorithm, which could produce erroneously high [Chl *a*] values. These issues will be discussed further in section 4.4.3.

The blended product primarily replicates the frequency distribution of Algal2 in the northern KZN bight (figure 4.32a), most likely due to the high persistence of OWT7 and Case 2 type waters in this region, where the Algal1 may be flagged. Over the southern KZN bight, the blend shows lower frequency between approximately 0.3 and 1 mg m^{-3}; this may represent a concentration range where OC4E could have been applied instead of Algal1 and Algal2. The blended product demonstrates a broader range of application in this subregion, with improved capability to resolve [Chl *a*] between approximately 1 and 10 mg m^{-3} compared to Algal1 and Algal2. The blended product thus demonstrates a similar application range than the standard satellite products with the improved capability of resolving [Chl *a*] across a wider range of water types.

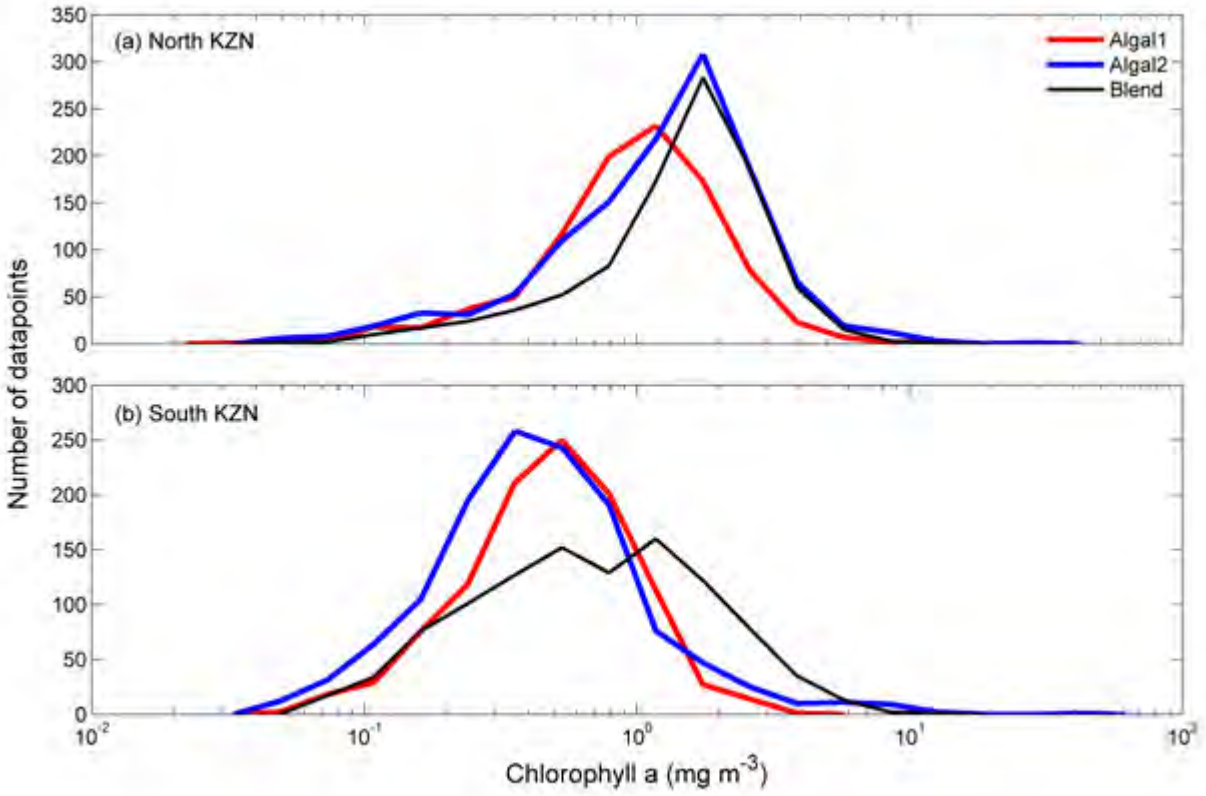


Figure 4.32: Histograms of the time series of daily [Chl *a*] products (as shown in figure 4.31) for the two east coast subregions (as detailed in figure 4.30). The products shown include Algal1 (red), Algal2 (blue) and the blended product (black).

4.3.6 [Chl *a*] climatologies

The seasonal cycle of the blended [Chl *a*] product for all the extracted subregions can be seen in figure 4.33; these plots represent temporal (monthly) and spatial (over the area of each subregion) averages of the blended [Chl *a*] for the entire MERIS mission which can be compared to other previous regional climatological and/or seasonal studies.

Figure 4.33a shows the seasonal cycle for the west coast subregions. The shapes are generally sinusoidal with summer peaks and winter troughs corresponding to the upwelling season of the southern Benguela. Overall the entire St Helena Bay subregion has a lower mean and variability than the two smaller coastal subregions, since it includes the highly variable coastal zone as well as low to medium biomass offshore waters. The amplitude and variability is highest for the Elands Bay region, which may result from the recurrent onshore and offshore movement and dynamic retention of phytoplankton blooms during the relaxation and upwelling phases of the upwelling season (Pitcher and Nelson, 2006). Studies have suggested that mean nearshore current flow in St Helena Bay is generally northward during spring and summer when upwelling winds are strongest, but southward towards the end of the upwelling season when there tends to be prolonged periods of relaxation of equatorward winds (Pitcher and Nelson, 2006; Fawcett et al., 2008); resultantly the Columbine subregion (in the south of St Helena Bay), a known bloom retention zone (Weeks et al., 2006), shows a peak later in the upwelling season than observed at Elands Bay when these phytoplankton blooms may be advected southwards by the near-shore currents.

The [Chl *a*] climatologies for the subregions of the inshore zone of the south coast are given in figure 4.33b. These areas represent the regions where the highest surface [Chl *a*] generally occur (Shannon et al., 1984; Demarcq et al., 2003), as previously suggested by the high persistence of the moderate to high biomass OWT5 in this region (figure 4.8). Although these subregions have lower average concentrations compared to those of the west coast, they show similar seasonal trends of winter minimum and April peak [Chl *a*] due to the prevalence of coastal wind-induced upwelling during summer (Schumann et al., 1982). The biological activity in the Algoa Bay subregion may also be influenced by nutrient enrichment from recurrent upwelling near Port Alfred (Walker, 1986; Lutjeharms, 2006). Winter conditions over the Agulhas Bank tend to show low [Chl *a*] that are evenly distributed in a well-mixed water column (Probyn et al., 1994); similar mean concentrations are observed for the coastal and shelf sites. A secondary peak in the Goukamma subregion in October corresponds to the initiation of the spring bloom on the

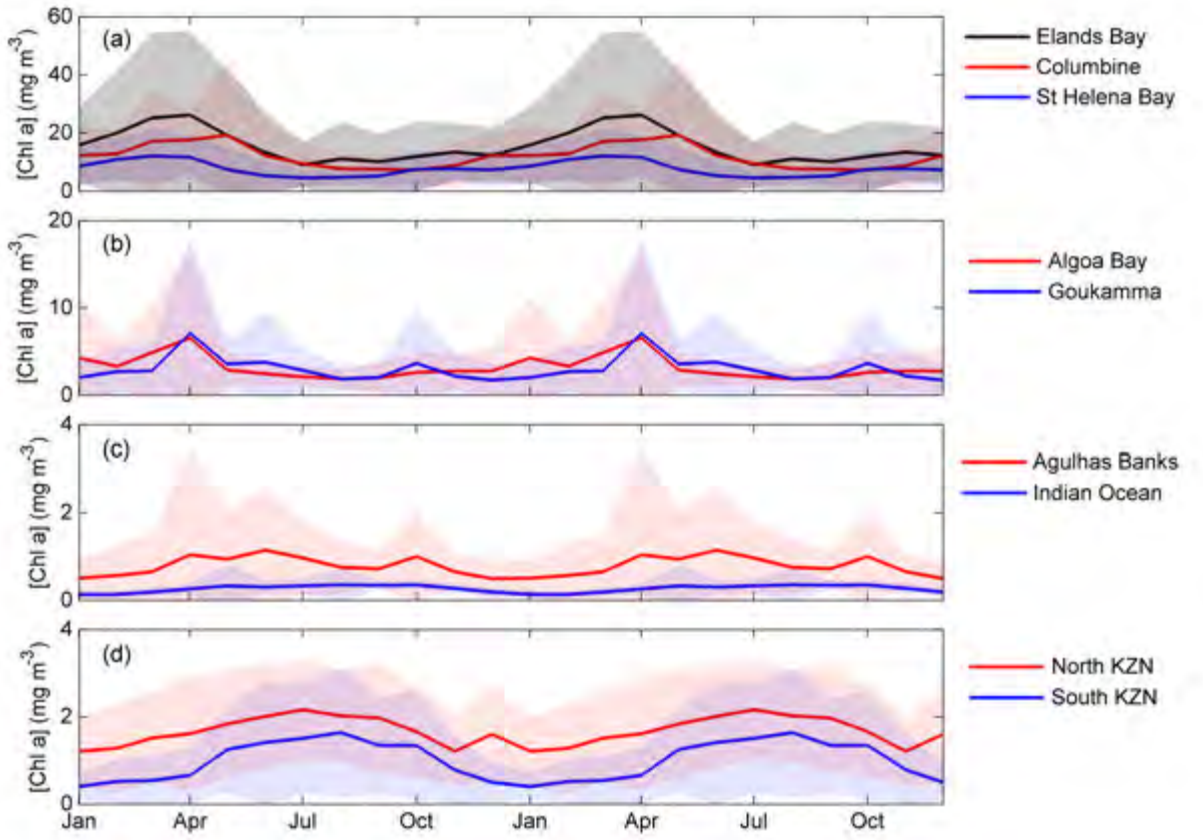


Figure 4.33: The monthly climatology of the blended Chl *a* product for each of the extracted regions. The data are grouped according to [Chl *a*] to increase visibility. The lines represent the monthly mean whilst shaded areas show the standard deviation. Note the changes of scale in the panels.

Agulhas Bank (Mitchell-Innes et al., 1999); the April and October chorophyll maxima were also observed in MODIS data (Jarre et al., 2015) in this region. The high intra-seasonal variability in [Chl *a*] along the south coast and Agulhas Bank is introduced through wind-driven upwelling and intrusions of Agulhas Current waters and the rapid response of the phytoplankton to this dynamic physical environment (Probyn et al., 1994).

In contrast to the results shown for the coastal zone, the offshore regions of the south coast (figure 4.33c) show a general trend of summer troughs and autumn and spring peaks. The Agulhas Bank is exposed to intense thermal stratification during summer, resulting in the formation of deep subsurface chlorophyll maxima (Probyn et al., 1994) whilst surface [Chl *a*] is generally low due to nutrient limitation in a stable water column (Brown, 1992; Mitchell-Innes et al., 1999). Demarcq et al. (2003) reported similar trends in surface [Chl *a*] from SeaWiFS climatologies over the central Agulhas Bank, with lowest and highest [Chl *a*] in January and June respectively. Brown (1992) reported chlorophyll maxima

during spring and autumn off the south coast typical of temperate continental shelf dynamics. Similar to the Goukamma region, the secondary peak in October corresponds to the initiation of the spring bloom on the Agulhas Bank (Mitchell-Innes et al., 1999); nutrient-rich water from coastal upwelling may also be advected over the shelf (Walker, 1986; Lutjeharms, 2006), influencing the phytoplankton productivity and variability on the Agulhas Bank (Lutjeharms et al., 2000a; Lutjeharms, 2007). The offshore Indian Ocean subregion represents the seasonal signal of oligotrophic oceanic waters. Similar seasonal trends have been observed in the southwest Indian ocean and Agulhas Current systems (Machu and Garçon, 2001; Demarcq et al., 2003; Lutjeharms, 2006; Jaquemet et al., 2007); Machu and Garçon (2001) suggested that the strengthening of the wind stress over the subtropical gyre in August could deepen the mixed-layer depth, which may in turn increase new production and phytoplankton stock.

The subregions of the KwaZulu-Natal Bight (panel d) both show a similar sinusoidal shape as the offshore subregions from the south coast; although the mean [Chl *a*] is elevated on the shelf compared to the oceanic environment (panel c), the seasonal trends reflect the physical characteristics and environmental conditions of the southwest Indian ocean and Agulhas Current. Lamont and Barlow (2015) showed the presence of subsurface maxima in summer (February) over the bight, which may explain the lower surface concentrations observed through satellite ocean colour methods during these months. The average [Chl *a*] for the northern bight is consistently between 0.5 and 1 mg m^{-3} higher than for the southern bight; Meyer et al. (2002) noted increased nutrient concentrations and [Chl *a*] in the northern bight and attributed it to the close proximity to the St Lucia upwelling cell. A recurrent lee eddy off Durban can advect nutrient-poor surface water from the Agulhas Current onto the southern bight (Lutjeharms et al., 2000a; Meyer et al., 2002), which may also lower average [Chl *a*] in this region. The northern bight has a relative increase in the average [Chl *a*] during December, possibly due to increased nutrient availability from riverine influxes at the beginning of summer; although some studies have suggested that terrigenous inputs of nutrients could promote productivity and increase phytoplankton biomass over the central bight (Ryther et al., 1966 and Burchall, 1968 cited in Lamont and Barlow, 2015), others have noted that the increased turbidity from riverine silt could potentially inhibit production through light limitation (Carter and Schleyer, 1988).

4.4 Discussion

4.4.1 Utility of the present set of OWT classes in the South African coastal sub-regions

This thesis provides the first optical characterization for South African coastal waters from OWTs developed within the region. The regional persistence of the dominant OWTs around South Africa reveals similar patterns as those observed by Mélin and Vantrepotte (2015), who found that the region to the east of the Agulhas Bank was dominated by OWTs that are generally associated with oligotrophic to mesotrophic conditions well suited to the use of Case 1 models, whilst the regions over the continental shelf of the west and south coasts were mostly represented by OWTs where standard Case 1 models might fail. The persistence of the current set of OWTs can be used indicatively to provide information of regional algorithm requirements; OWTs 1 to 4 show areas where blue-green algorithms can be used optimally, OWTs 5 and 6 represent regions where red-NIR algorithms may be required, whilst OWTs 7 to 11 show areas which necessitate more complex (and potentially regionally-specific) algorithms able to deal with Case 2 waters. From the OWT persistence maps it can be seen that SA coastal waters are primarily dominated by the moderate reflectance classes, i.e. where phytoplankton IOPs have the greatest impact on the reflectance magnitude and shape. Generally only small strips of nearshore waters were affected by the elevated reflectance classes. The northern KZN Bight showed different optical conditions prevailing compared to the rest of the coast, with the farthest offshore extent and highest persistence of the elevated reflectance classes. A hard classification switching algorithm has previously been suggested for the KwaZulu-Natal bight with Algal1 switching to Algal2 when the Algal1 quality flag was raised (Smith et al., 2013); although the new OWT framework still applies similar algorithms, it now offers an improvement on the original scheme by blending of the [Chl *a*] products through OWT membership functions whilst avoiding potential spatial discontinuities imposed by the original hard binary classification.

One of the anticipated applications for the current set of OWTs is the ability to classify high biomass phytoplankton blooms spectrally instead of using derived products (e.g. [Chl *a*]). Harmful Algal Blooms (HABs), which occur most frequently from January to May in the latter half of the upwelling season (Pitcher and Calder, 2000) in the southern Benguela, are usually attributed to dinoflagellate species (Pitcher and Weeks, 2006); these blooms can negatively impact commercial and recreational activities in the region through causing

hypoxia following the decay of a large bloom (Pitcher et al., 1998) or by producing toxic compounds which can contaminate shellfish and/or cause mortalities of various marine organisms (Pitcher and Calder, 2000; Probyn et al., 2000). Previous studies of bloom detection metrics have used [Chl *a*] thresholds to classify blooms in the southern Benguela (Evers-King, 2014), and have primarily focussed on the upwelling season. Results from the current study suggests that dinoflagellate-type assemblages classified as OWT6 are present in St Helena bay more often than previously reported. OWT6 shows the highest persistence during the summer months, but also maintains a steady presence in the bay throughout the year; however, the OWT method is only qualitative and the blended [Chl *a*] climatology for the region still shows a clear seasonal signal with an autumn peak and winter trough. With the provision of a regionally-appropriate atmospheric correction algorithm the OWT classification method (specifically utilizing OWT6) could be applied to OLCI data to potentially form part of an operational HAB monitoring system.

The elevated reflectance classes may be applicable for studies focussing on the ecological impact of turbidity, which is known to affect fish distributions in estuaries along the east coast (Cyrus and Blaber, 1987; Harris et al., 2001). During a ten year study of the movements of sardine shoals during the KZN sardine run O'Donoghue et al. (2010) found that the presence of turbid waters (among other factors) was found to have a negative impact on sardine presence, which can have regional socio-economic and ecological impacts (Hutchings et al., 2010). Utilizing existing classes or regionally-parameterized OWTs, together with high resolution satellite data such as OLCI FR, will prove beneficial to these types of studies.

The OWT framework provides the capability of exploiting a wide range of operational algorithms by selecting and constraining application to optimized bio-optical ranges. Results in chapter 3 demonstrated that the current OWT framework and algorithm blending procedure can provide retrievals with a median absolute relative error of 45.5% over a range of [Chl *a*] spanning almost four orders of magnitude. The blended [Chl *a*] product provides an improvement on standard satellite products for the west and south coast since it can capture the extent of high biomass phytoplankton blooms, whilst showing improved capability to resolve low phytoplankton biomass compared to currently available regional products. Even though the OWTs used in this study are optimized for the west coast, the OWT blending procedure has been shown to perform adequately along the south and east coasts when atmospheric correction was successful. This study reinforces the concept that the algorithm blending approach provides improved results compared to one algorithm

applied over a wide dynamic range of environmental conditions (Moore et al., 2014).

4.4.2 Limitations of the current set of optical water types and the classification approach

As demonstrated previously the current OWT algorithm is regionally biased to the optical conditions in the southern Benguela in terms of the bulk of the *in situ* data and the extracted satellite data of the clustering database. Apart from OWTs 7 and 8, which both contain a high number of spectra consisting of mostly MERIS data, the other three elevated reflectance classes showed very low persistence (or none at all) around the rest of the coastline; these OWTs consisted of primarily synthetic data and MERIS data from the southern Benguela. Moore et al. (2001) suggested that the suspended mineral sediments play a large role in determining the shape of the reflectance spectra and that these classes are likely region-specific; thus it is possible that the water types optimized for the southern Benguela would not necessarily be found elsewhere along the SA coast. A detailed regional study focussing on *in situ* radiometric data collection and/or satellite data extraction in regions known to be affected by fluvial sediments may be required for a more comprehensive optical characterization of the different elevated reflectance classes. Such a study can be guided by the classification results from this thesis.

One of the implied constraints of the clustering techniques applied in this thesis is the moderate versus elevated reflectance database division; resultantly OWT7 is comparable to OWT5 in that it represents groupings of the lowest magnitude reflectances of the moderate and elevated reflectance databases. OWT7 clearly contains a variety of spectral shapes that could potentially form separate (albeit small) clusters if a more holistic approach were to be used during the clustering process. For the purposes of the current body of work the technique appears to be valid, since it provides a type of investigative separator for phytoplankton dominated waters (whether oligotrophic or eutrophic) and turbid waters with varying degrees of biomass. However, this bipartite approach may not be advised if regional algorithm development was the desired application, in which case the normalization approach may be more appropriate (Vantrepotte et al., 2012; Mélin and Vantrepotte, 2015).

As seen in the results the dominance of OWT1 was very low. It is unlikely that the very low biomass waters which OWT1 represents were not present in the MERIS imagery, especially with the known near-oligotrophic nature of the offshore east coast region (Smith

et al., 2013); however, since the class consists of primarily synthetic data, it is possible that it is not optimally paramaterized for these areas, or even for the offshore southern Benguela region. Mélin and Vantrepotte (2015) demonstrated that optical variability exists even in very oligotrophic regions, which can generally be attributed to the regional influence from external sources such as terrigenous dust or CDOM (e.g. Claustre et al., 2002; Morel and Gentili, 2009). For operational use, OWTs 1 and 2 could potentially be combined due to the similarities in their spectral shapes and determinative modes of variance; the resultant expansion in cluster size and covariance matrix inclusivity could potentially lead to increased classification success. These results also show that the use of synthetic data should be carefully considered, since such data may not represent conditions typically found in regional water bodies.

The switching of algorithms that occur within OWT5 is perhaps not ideal as it could still, on occasion, produce hard boundaries; although this may not be apparent in the [Chl *a*] images, discontinuities can be distinguished in histograms as demonstrated in figure 4.23a. However, this approach has provided a practical compromise for optimized algorithm application with the current OWTs. Further parameterization of the absorption-dominated water types may be needed, perhaps using reflectance ratios or re-clustering this highly absorbing class into subclusters. However, the optimal assessment of these improvements would always depend on the ability of the atmospheric correction to determine accurate water-leaving reflectance, particularly in the red-NIR spectral region which is often used to resolve high biomass [Chl *a*]. Any further regional studies should target the quantitative assessment and improvement of atmospheric correction as a priority.

4.4.3 Impacts of the satellite data quality and coverage on the stated key research questions of the study

The coverage and quality of MERIS data has been sufficient to enable the assessment of the general dominance and persistence of the OWTs produced in this study. The regions associated with dominance of the various moderate reflectance classes are consistent with known trophic regions in the literature, whilst the persistence of the elevated reflectance classes offer new insight into the frequent occurrence of moderate to highly scattering waters (e.g. in the northern KZN Bight, and the near-shore regions of St Helena Bay) that may require Case 2 or regional inversion algorithms. The representativeness of the OWTs are generally higher in the southern Benguela, since the OWT classes were constructed around *in situ* and extracted satellite data from this region. The OWTs were optimized for MERIS application with reflectance data from the bright pixel atmospheric correction (BPAC); however, different atmospheric correction techniques may produce spectra with shapes and magnitudes that are not consistent with these OWT classes, resulting in lower classification success. This may be the case for the east coast and offshore regions, where BPAC is not consistently applied. BPAC is occasionally applied over the coastal waters of the northern KZN Bight, where the average membership sums are slightly higher than the rest of the east coast near-shore waters. Increased persistence of average membership sums (over the MERIS time series) may potentially be achieved by applying the PCD1_13 flag before OWT classification; future OWT classification applications should consider applying the PCD1_13 flag (or similar atmospheric correction failure indicators) routinely to properly constrain input data when assessing OWT persistence and blended [Chl a] products.

There is a strip of decreased average membership sums and low membership in the near-shore waters along almost the entire country that could potentially be the result of the adjacency effect; this effect occurs when light from land is scattered into the sensor's field of view. The adjacency effect often produces an increase in at-sensor radiance in the NIR (Santer and Vidot, 2004), which may be mistaken for increased aerosol optical thickness during atmospheric correction procedures, and can resultantly lead to over-correction of shorter wavelengths (Binding et al., 2011). The improved contrast between ocean and land (ICOL, Santer, 2010) processor, a method for the correction of land adjacency correction specifically designed for MERIS, has been shown to have a slight improvement to reflectance errors along the east coast (Smith et al., 2013); however, it has also produced some unusual spectral shapes in South African dams and coastal waters (Matthews et al.,

2012). New techniques for the correction of adjacency effects (e.g. Sterckx et al., 2015) or atmospheric corrections over optically complex waters (e.g. Mao et al., 2013; Singh and Shanmugam, 2014) may need to be considered for future OWT classification endeavours in South African coastal waters.

The ability of the classification framework to identify OWTs and subsequently apply and blend Chl *a* algorithms at the event scale was demonstrated for the case studies in sections 4.3.3 to 4.3.5. The ideal circumstances for the classification framework to provide reliable [Chl *a*] are under conditions of high total membership sums, since this would imply that the surface waters in the image are well represented by the OWT classes and that the applied Chl *a* algorithms are generally appropriate. However, the use of weighted fuzzy memberships implies that any successful pixel classification, even with very low membership sums, would result in algorithm application. This results in the retrieval of a [Chl *a*] product even where OWTs are not optimally represented with high total membership sums, as was the case along the east coast. Quality assurance measures are set in place to avoid the application of algorithms to possibly erroneous reflectance data by the application of the PCD1_13 flag to the [Chl *a*] product; an algorithm is also not applied if a pixel is left unclassified, i.e. if the spectral shape of the pixel does not conform in any quantitative way to the available OWT classes. It was demonstrated that, under conditions of acceptable atmospheric correction, the OWT classification and blending framework can successfully retrieve a larger range of [Chl *a*] than standard and regional satellite products across a dynamic range of optical conditions. A low membership threshold could also be applied to indicate poor retrieval of OWTs and to flag out [Chl *a*] estimates with higher levels of uncertainty.

With or without the application of the PCD1_13 flag, there were generally less data available during summer and spring due to the high persistence of cloud and uncorrected sun glint (figure 4.6). Although these quality flags are vital to ensure a low uncertainty for the output products, their application can limit the number of data available, potentially introducing sources of uncertainty in the representativeness of the remaining data. In data-sparse regions event-scale phytoplankton blooms may be missed entirely (Mitchell et al., 1991), or conversely may disproportionately bias the spatial or temporal [Chl *a*] averages (Gregg and Casey, 2007; Rivas et al., 2006). The near-shore zone of the east coast suffers particularly from low data availability since valid pixels can be limited to less than 10% (in some cases less than 10 pixels) per season over a 10 year period. Time series data from this region should be interpreted with caution. Despite the power of satellite

remote sensing, clearly it has limits in cloudy regions and therefore *in situ* measurements may in such cases take precedence.

The ability of the OWT classification to assign the appropriate algorithm per pixel rests upon the ability of the atmospheric correction algorithm to retrieve the water-leaving reflectance spectrum. Although OWT6 has shown good synoptic coherence with high biomass dinoflagellate blooms, the atmospheric correction over these highly absorbing waters is still challenging and failure can result in low membership sums, improper classification, or unclassified pixels. In some cases the most severe parts of the blooms may be flagged by PCD1_13 or left unclassified if atmospheric correction problems produce negative spectra; this may mean that some of the pixels representing the highest biomass regions and potentially harmful blooms may not be accounted for in the summer and autumn months. Improper atmospheric correction has the most notable effect on the classification potential of OWT5: firstly, the relatively low reflectance of this class may be subject to a significant level of noise (Mélin and Vantrepotte, 2015); secondly, it is possible that an over-correction in the blue-green (from the adjacency effect, as mentioned above) could produce spectra with erroneously low reflectance that may be mistakenly classified into this OWT. This could potentially explain the occurrence of OWT5 along the relatively unproductive east coast (up to 20% in figure 4.8e), where both atmospheric correction and adjacency are potential problems. Since this class necessitates an algorithm switch due to the wide range of [Chl *a*] that is represented, there is the possibility that a red-NIR algorithm may be improperly applied, which can produce a lower quality [Chl *a*] product. A possible solution could be to forgo the algorithm switch along the east coast (i.e. only use OC4E for OWT5), where historical *in situ* surface [Chl *a*] measurements have been found to be primarily below the switching threshold (Carter and Schleyer, 1988; Barlow et al., 2008, 2013; Smith et al., 2013; Barlow et al., 2015).

4.4.4 Conclusions and recommendations

This chapter demonstrated the satellite application of the newly defined set of OWTs through the characterization of the spatial bio-optical variability in South African waters in addition to successful application and blending of class-appropriate algorithms across wide ranges of [Chl *a*] and water types in different regions. The need and utility of red-NIR spectral bands for capturing the extent of phytoplankton blooms along the west and south coast has been clearly demonstrated. As the MERIS follow-up mission OLCI provides the obvious choice to continue the OWT classification in these regions; due to the similar spectral resolution the OWT and algorithms defined in this study can be applied to OLCI with minimal changes, whilst error characterization can follow once match-up data has been acquired. OLCI in constellation onboard three satellites will offer effective daily revisit times over South Africa (Donlon et al., 2012) with diminished constraints of regional glint and low coverage. The OWT persistence and low membership maps can also be used to guide future validation efforts and algorithm development for these and in next generation satellites.

The success of the OWT classification method relies heavily on suitable atmospheric correction which can preserve the spectral shapes and amplitudes of the water leaving reflectance. The persistence of low memberships sums with coincident data quality flags have indicated that there is a clear need for improved atmospheric correction and possibly adjacency effect correction, particularly in the coastal water along the south and east coasts. Resultantly there is generally low data availability and quality from MERIS for the east coast. Due to the absence of high phytoplankton biomass in this region there is less dependency on the red-NIR spectral region for the classification and algorithm application process. OWT parameterization and persistence studies may consider utilizing MODIS for this region, which offers various atmospheric correction schemes (see examples in Goyens et al., 2013) and suites of derived products such as [Chl *a*] (e.g. O'Reilly et al., 2000; Carder et al., 2004) and turbidity (e.g. Wang et al., 2009); there is scope to parameterize existing semi-analytical algorithms for the region (e.g. Magnuson et al., 2004), whilst globally-defined OWTs are available for application to MODIS data (Moore et al., 2009). Two versions of MODIS (Aqua and Terra) are currently or have been in operation which have different overpass times (Terra in the morning and Aqua in the afternoon); although MODIS on Terra has had radiometric degradation issues (Franz et al., 2008), various cross-calibration techniques have been developed to improve the radiometric products (Kwiatkowska et al., 2008; Meister and Franz, 2011; Meister et al., 2014). The utility of data from both sensors could potentially increase the temporal coverage for the east coast

and potentially lead to the availability of 16 years of ocean colour data for the region. There is thus considerable benefit to realising the full volume of MODIS sensors for the region.

The low persistence of many of the elevated reflectance classes (OWTs 9 to 11) point to the lack of dedicated optical characterization of the turbid waters along the south and east coasts. These areas could benefit from increased *in situ* data collection, or regional extraction of satellite spectra under known high flow conditions in order to characterize the spectral shapes and magnitudes associated with the receiving waters of regional rivers (e.g. Thukela River). In this study the Algal2 product was used as a convenient available product for elevated reflectance classes in MERIS data, but no substantial regional validation has been performed for the Algal2 product under sediment-dominated, turbid, or highly scattering conditions. Improved availability of validation data for moderate to highly scattering waters will lead to more optimized algorithms and parameterizations for the elevated reflectance classes.

The OWTs applied in this study are the result of a specific clustering method (FCM) on specific datasets containing *in situ*, synthetic and extracted satellite reflectance spectra; an optimal set of OWTs were then determined prioritising the ecological relevance determined primarily by [Chl *a*] applications. Data from any of these sources can be removed or added depending on the desired classification applications. Although improved classification success (through sensor optimization) and a more comprehensive optical characterization of the water types around South Africa could be achieved by representing all coastal regions and seasons equally in a satellite data extraction process (as per methods in Mélin and Vantrepotte, 2015), synthetic and *in situ* data offer spectral and bio-optical information at a lower uncertainty than satellite data which can be utilized for algorithm development and error characterization. Moore et al. (2009) suggested that as more data are added, the OWT characteristics would evolve and eventually converge to form a stable set. The OWT framework is amenable to change and more *in situ* data can be added as they become available, whilst newly-validated or regionally-parameterized algorithms may be substituted during the algorithm blending process.

Chapter 5

Conclusions and recommendations for further work

5.1 Conclusions

This study presents the first dedicated characterization of the optical waters types of the coastal waters of South Africa. Chapters 2 and 3 demonstrate a systematic clustering database expansion and modification process which combines reflectance data from *in situ*, synthetic and satellite sources to determine the most appropriate set of regional clusters with regard to ecologically-focussed application of [Chl *a*] products. A clustering database separation, based on the amplitude of reflectance at 490 and 560 nm, facilitates the clustering process by differentiating between spectra representing phytoplankton-dominated and moderate to highly scattering waters. The final set of clusters consists of six phytoplankton-dominated OWTs representing a range from near-oligotrophic oceanic waters to shelf-scale eutrophic phytoplankton bloom waters, as well as a selection of five moderate to highly scattering OWTs primarily representative of near-shore Case 2 type conditions.

The performance of a selection of known and well-documented Chl *a* algorithms was assessed per OWT and assigned based on lowest RMSE. It was found that the [Chl *a*] of two of the OWTs, representing moderate to very high phytoplankton biomass, were best resolved with the combined use of a blue-green algorithm switching to a red-NIR algorithm above a certain threshold; thus a OC4E-G2B at 10 mg m^{-3} and OC4E-G3B at 15 mg m^{-3} was implemented for OWTs 5 and 6, respectively. The assigned algorithms were implemented as part of a classification and algorithm blending framework that produced improved RMSE and MARE results compared to any single satellite algorithm across a

wide range of water types. The classification and blending approach was able to retrieve [Chl *a*] with total uncertainty of 45.5% and a positive bias of 0.087 of from a global satellite validation dataset spanning almost four orders of magnitude. These results clearly show that selecting and constraining algorithm application to optimized bio-optical ranges produces lower uncertainty products than achievable by any single algorithm for South African waters.

In chapter 4, the eleven OWTs produced in this study were applied to ten years of MERIS data and demonstrated the value of the classification approach in characterizing, for the first time, the phenology and spatial bio-optical variability found in South African coastal waters. The high persistence of moderate to high phytoplankton biomass, as well as the occasional occurrence of very high biomass dinoflagellate bloom events, in the productive coastal waters of the west and south coast, necessitates the availability and utility of red-NIR spectral bands for quantitative [Chl a] retrievals from satellite radiometry. The results from this thesis also serve as a user guide for ocean colour utility in the region by demonstrating the long term bio-optical variability within South African coastal waters, in addition to providing the realised usable satellite product observation frequency, at least for MERIS, which has important implications with regard to ocean colour applications in the highly dynamic coastal environment.

This study highlighted the effects that discrepancies between *in situ* and satellite reflectance data can have on classification ability, and the importance of understanding the sensor-specific quantitative effects that the atmospheric correction may have on classification ability. Although the total average membership sums decreased towards the east coast, the classification framework was able to, under conditions of appropriate atmospheric correction, effectively classify, assign and blend class appropriate algorithms on a per pixel basis into synoptically coherent class maps and [Chl *a*] products. The results from this study confirm the hypothesis that the satellite-retrieved [Chl *a*] product for South African coastal waters can be improved by the application of a regional classification framework that applies and blends appropriately constrained class-appropriate algorithms on a per-pixel basis. The resulting blended product provides improved application ability compared to currently available standard and regional satellite products by providing an optimized [Chl *a*] across a higher dynamic range of concentrations and bio-optical conditions.

5.2 Considerations for further work

- There is a clear need for further collection of *in situ* data from South African coastal regions that include sensor-independent hyperspectral radiometric data, with co-incident biogeochemical parameters and IOPs. These data should include more moderate to highly scattering conditions in order to better characterize the various spectral shapes, magnitudes, and IOP ranges associated with these types of waters; these data can be used to derive optimized algorithms, or properly parameterize existing Case 2 algorithms, per OWT.

- The maps of persistent low class membership resulting from the classification approach applied in this thesis point out regions of potentially missing optical water types and/or problems with the atmospheric correction. Future *in situ* data collection and validation, as well as atmospheric correction improvement efforts, can be guided by these results to best focus research efforts.

- There is a clear need for improved atmospheric correction ability along the east coast and other coastal regions of South Africa. Future efforts should consider including adjacency corrections, whilst focussing on the quantitative assessment of the correction ability to maintain the spectral shape and magnitude of the water-leaving reflectance. Some consideration should also be given to avoiding aerosol correction issues by classifying directly on Rayleigh-corrected reflectance; this would expand the classification approach to allow water-type based decisions higher up in the processing chain, which would prove particularly useful for inland and near-coastal waters.

- The moderate to very high biomass classes could potentially be re-parameterized in order to minimize or negate the need for a switch in the algorithm application within an OWT. This would however still be dependent on the ability of the atmospheric correction to retrieve accurate reflectance data, as poorly corrected reflectance are still at risk for erroneous classification.

- Algal2 was assigned to the elevated reflectance classes as a pragmatic solution to dealing with Case 2 type waters. However, more work is required in assessing the performance of, and potentially re-paramaterizing, a wider range of Case 2 algorithms capable of retrieving low uncertainty [Chl *a*] products within the moderate to highly scattering coastal waters of South Africa.

- Development of shallow water classes should be considered to enable improved differentiation between the effects of bottom reflectance and backscattering from the particles within the water column; regions for potential application could include the south-eastern shores of St Helena Bay or in the lagoon of Saldanha Bay. These classes could either be utilized within the algorithm application framework if appropriate algorithms exist, or as a data quality flag. This is a difficult task, and is linked with the need for *in situ* data collection, as pointed out above.

- The ocean colour community has expressed the need for per pixel uncertainty products for ocean colour data products. Future OWT classification applications should focus on determining improved uncertainty estimates per OWT and algorithm in order to have the capability of providing optimized blended uncertainty products. Further thought should be given to the actual uncertainty application, e.g. how to use such data for integrated modelling or phenological applications.

- The OWTs described in this thesis can be applied, with minimal adaptation, to OLCI reflectance data. This will enable operational implementation of the OWT classification framework for applications such as harmful algal bloom monitoring and coastal water quality management. This should be considered a very high priority for operational implementation. Making these data and uncertainty products available in a public space for user uptake would be the first step.

- This OWTs presented in this thesis, as well as the class-specific algorithms assigned to the OWTs, are amenable to change. Newly collected *in situ* data can be incorporated into the classification framework through re-clustering, whilst newly defined algorithms that have not yet been tested within the region can be assigned per OWT as they become available. Operational implementation of OWT products should therefore consider the OWT framework as dynamic, i.e. periodic improvement and reprocessing will be required.

- Although this thesis uses the above-water remote sensing reflectance ($R_{rs}(0+)$), other bodies of OWT classification work have often used the in-water version ($R_{rs}(0-)$) (e.g. Moore et al., 2001, 2009, 2012, 2014). Since the $R_{rs}(0-)$ values are larger (resulting in larger means and covariances) the use of this variable may improve the classification success. Future studies should aim to develop means and covariance matrices, and to classify satellite imagery, using the $R_{rs}(0-)$ variable.

- The FCM clustering algorithm and the satellite image classification scheme uses the Euclidean distance and the Mahalanobis distance metrics respectively; how-

ever, these two schemes may produce slightly different cluster structures. The re-classification of the initial cluster points with the Mahalanobis distance metric may re-assign points to different clusters; this method may produce clusters that are more appropriate for satellite image applications, since the modified clusters are created in a similar way to how the satellite data are treated during the classification process.

References

Ahn, Y. (1999). *Proprietes optiques des particules biologiques et minerales*. PhD thesis, Pierre and Marie Curie University, Paris, France.

Aiken, J. and Moore, G. (2000). ATBD Case 2 s bright pixel atmospheric correction. *Center for Coastal & Marine Sciences, Plymouth Marine Laboratory, UK, Rep, PO-TN-MEL-GS*, 4(0005).

Ainsworth, E. and Jones, I. (1999). Radiance spectra classification from the Ocean Color and Temperature Scanner on ADEOS. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1645–1656.

Alimonte, D. D., Mélin, F., Zibordi, G., and Berthon, J.-F. (2003). Use of the novelty detection technique to identify the range of applicability of empirical ocean color algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2833–2843.

Alvain, S., Loisel, H., and Dessailly, D. (2012). Theoretical analysis of ocean color radiances anomalies and implications for phytoplankton groups detection in case 1 waters. *Optics express*, 20(2):1070–1083.

Alvain, S., Moulin, C., Dandonneau, Y., and Bréon, F. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep Sea Research Part I: Oceanographic Research Papers*, 52(11):1989–2004.

Ambarwulan, W., Salama, M., Mannaerts, C., and Verhoef, W. (2011). Estimating specific inherent optical properties of tropical coastal waters using bio-optical model inversion and in situ measurements: case of the Berau estuary, East Kalimantan, Indonesia. *Hydrobiologia*, 658(1):197–211.

Antoine, D., D'Ortenzio, F., Hooker, S. B., Bécu, G., Gentili, B., Tailliez, D., and Scott, A. J. (2008). Assessment of uncertainty in the ocean reflectance determined by three satellite ocean color sensors (MERIS, SeaWiFS and MODIS-A) at an offshore site in the Mediterranean Sea (BOUSSOLE project). *Journal of Geophysical Research*, 113(C7).

Antoine, D. and Morel, A. (2005). MERIS ATBD 2.7. Atmospheric correction of the MERIS observations over ocean Case 1 waters. Issue 5, Revision 0.

Arnone, R. A., Wood, A. M., and Gould, R. (2004). SCIENCE BOX: The Evolution of Optical Water Mass Classification. *Oceanography*, 17(2):14–15.

Bao, Y., Tian, Q., and Chen, M. (2015). A Weighted Algorithm Based on Normalized Mutual Information for Estimating the Chlorophyll-a Concentration in Inland Waters Using Geostationary Ocean Color Imager (GOCI) Data. *Remote Sensing*, 7(9):11731–11752.

Barker, K., Mazeran, C., Lerebourg, C., Bouvet, M., Antoine, D., Ondrusek, M., Zibordi, G., and Lavender, S. (2008). MERMAID: The MERis MAtchup In-situ Database. In *Proceedings of the 2nd MERIS/(A) ATSR workshop ESA SP*, volume 666.

Barlow, R. (1982). Phytoplankton ecology in the southern Benguela current. I. Biochemical composition. *Journal of Experimental Marine Biology and Ecology*, 63(3):209–227.

Barlow, R., Gibberd, M., Lamont, T., Aiken, J., and Holligan, P. (2016). Chemotaxonomic phytoplankton patterns on the eastern boundary of the Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 111:73–78.

Barlow, R., Kyewalyanga, M., Sessions, H., van den Berg, M., and Morris, T. (2008). Phytoplankton pigments, functional types, and absorption properties in the Delagoa and Natal Bights of the Agulhas ecosystem. *Estuarine, Coastal and Shelf Science*, 80(2):201–211.

Barlow, R., Lamont, T., Britz, K., and Sessions, H. (2013). Mechanisms of phytoplankton adaptation to environmental variability in a shelf ecosystem. *Estuarine, Coastal and Shelf Science*, 133:45–57.

Barlow, R., Lamont, T., Gibberd, M., van den Berg, M., and Britz, K. (2015). Chemotaxonomic investigation of phytoplankton in the shelf ecosystem of the KwaZulu-Natal Bight, South Africa. *African Journal of Marine Science*, 37(4):467–484.

Barlow, R., Lamont, T., Kyewalyanga, M., Sessions, H., and Morris, T. (2010). Phytoplankton production and physiological adaptation on the southeastern shelf of the Agulhas ecosystem. *Continental Shelf Research*, 30(13):1472–1486.

Barlow, R., Sessions, H., Balarin, M., Weeks, S., Whittle, C., and Hutchings, L. (2005). Seasonal variation in phytoplankton in the southern Benguela: pigment indices and ocean colour. *African Journal of Marine Science*, 27(1):275–287.

Barlow, R., Sessions, H., Silulwane, N., Engel, H., Hooker, S., Aiken, J., Fishwick, J., Vicente, V., Morel, A., Chami, M., et al. (2003). BENCAL cruise report. *SeaWiFS Postlauch Technical Report Series*, 27:1–64.

Bastiaanssen, W. G. et al. (1998). *Remote sensing in water resources management: the state of the art.* International Water Management Institute.

Beckley, L. (1988). Spatial and temporal variability in sea temperature in Algoa Bay, South Africa. *South African Journal of Science*, 84(1):67–69.

Behrenfeld, M. J., Westberry, T. K., Boss, E. S., O'Malley, R. T., Siegel, D. A., Wiggert, J. D., Franz, B., McClain, C., Feldman, G., Doney, S. C., et al. (2009). Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences*, 6(5).

Benfield, S., Guzman, H., Mair, J., and Young, J. (2007). Mapping the distribution of coral reefs and associated sublittoral habitats in Pacific Panama: a comparison of optical satellite sensors and classification methodologies. *International Journal of Remote Sensing*, 28(22):5047–5070.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A., and Murtagh, R. F. (1996). Validity-guided (re) clustering with applications to image segmentation. *Fuzzy Systems, IEEE Transactions on*, 4(2):112–123.

Bernard, S., Balt, C., Pitcher, G., et al. (2005). The use of MERIS for harmful algal bloom monitoring in the southern Benguela. In *MERIS (A) ATSR Workshop 2005*, volume 597, page 26.

Bernard, S., Pitcher, G., Evers-King, H., Robertson, L., Matthews, M., Rabagliati, A., and Balt, C. (2014). Ocean Colour Remote Sensing of Harmful Algal Blooms in the Benguela System. In *Remote Sensing of the African Seas*, pages 185–203. Springer.

Bezdek, J. C. (1974). Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1(1):57–71.

Bezdek, J. C. (1975). Mathematical models for systematics and taxonomy. In *Proceedings of eighth international conference on numerical taxonomy*, volume 3, pages 143–166.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers.

Binding, C., Greenberg, T., Jerome, J., Bukata, R., and Letourneau, G. (2011). An assessment of MERIS algal products during an intense bloom in Lake of the Woods. *Journal of Plankton Research*, 33(5):793–806.

Blondeau-Patissier, D., Gower, J. F., Dekker, A. G., Phinn, S. R., and Brando, V. E. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in Oceanography*, 123:123–144.

Bosman, C., Uken, R., Leuci, R., Smith, A., and Sinclair, D. (2007). Shelf sediments off the Thukela River mouth: complex interaction between fluvial and oceanographic processes. *South African Journal of Science*, 103(11-12):490–492.

Bracher, a., Taylor, M. H., Taylor, B., Dinter, T., Röttgers, R., and Steinmetz, F. (2015). Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Ocean Science*, 11:139–158.

Bricaud, A., Morel, A., and Prieur, L. (1981). Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnology and oceanography*, 26(1):43–53.

Brockmann, C. (2006). Limitations of the application of the MERIS atmospheric correction. In *Proceedings of the Second Working Meeting on MERIS and AATSR Calibration and Geophysical Validation (MAVT-2006)*, pages 19–1.

Brown, C. W. and Yoder, J. A. (1994). Coccolithophorid blooms in the global ocean. *Journal of Geophysical Research: Oceans*, 99(C4):7467–7482.

Brown, P. (1992). Spatial and seasonal variation in chlorophyll distribution in the upper 30 m of the photic zone in the southern Benguela/Agulhas ecosystem. *South African Journal of Marine Science*, 12(1):515–525.

Brown, P., Painting, S., and Cochrane, K. (1991). Estimates of phytoplankton and bacterial biomass and production in the northern and southern Benguela ecosystems. *South African Journal of Marine Science*, 11(1):537–564.

Burchall, J. (1968). *Primary production studies in the Agulhas Current region off Natal-June, 1965*. South African Association for Marine Biological Research.

Campbell, J. W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research: Oceans (1978–2012)*, 100(C7):13237–13254.

Cannizzaro, J. P. and Carder, K. L. (2006a). Estimating chlorophyll a concentrations from remote-sensing reflectance in optically shallow waters. *Remote Sensing of Environment*, 101(1):13–24.

Cannizzaro, J. P. and Carder, K. L. (2006b). Estimating chlorophyll *a* concentrations from remote-sensing reflectance in optically shallow waters. *Remote Sensing of Environment*, 101(1):13–24.

Carder, K., Chen, F., Cannizzaro, J., Campbell, J., and Mitchell, B. (2004). Performance of the MODIS semi-analytical ocean color algorithm for chlorophyll-a. *Advances in Space Research*, 33(7):1152–1159.

Carder, K. L., Chen, F., Lee, Z., Hawes, S., and Kamykowski, D. (1999). Semianalytic Moderate-Resolution Imaging Spectrometer algorithms for chlorophyll a and absorption with bio-optical domains based on nitrate-depletion temperatures. *Journal of Geophysical Research: Oceans (1978–2012)*, 104(C3):5403–5421.

Carter, R. and d'Aubrey, J. (1988). Inorganic nutrients in Natal continental shelf waters. *Coastal Ocean Studies off Natal, South Africa*, pages 131–151.

Carter, R. and Schleyer, M. (1988). Plankton distributions in Natal coastal waters. *Coastal Ocean Studies off Natal, South Africa*, pages 152–177.

Carver, S. A., Siegel, D. A., and Mitchell, B. G. (1994). Variability in near-surface particulate absorption spectra: What can a satellite ocean color imager see? *Limnology and Oceanography*, 39(6):1349–1367.

Central Intelligence Agency (1977). *World Data Bank II: North America, South America, Europe, Africa, Asia*. http://doi.org/10.3886/ICPSR08376.v1 [2006-01-18]. Accessed: 2016-07-21.

Chen, X., Li, Y. S., Liu, Z., Yin, K., Li, Z., Wai, O. W., and King, B. (2004). Integration of multi-source data for water quality classification in the Pearl River estuary and its adjacent coastal waters of Hong Kong. *Continental Shelf Research*, 24(16):1827–1843.

Cipollini, P., Corsini, G., Diani, M., and Grasso, R. (2001). Retrieval of sea water optically active parameters from hyperspectral data by means of generalized radial basis function neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1508–1524.

Clark, D. K., Yarbrough, M. A., Feinholz, M., Flora, S., Broenkow, W., Kim, Y. S., Johnson, B. C., Brown, S. W., Yuen, M., and Mueller, J. L. (2003). MOBY, A Radiometric Buoy for Performance Monitoring and Vicarious Calibration of Satellite Ocean Color Sensors: Measurement and Data Analysis Protocols. In Mueller, J., Fargion, G., and McClain, C., editors, *Ocean Optics Protocols for Satellite Ocean Color Sensor validation, NASA Tech. Memo. 2003 - 211621/Rev4, vol. VI.* NASA GSFC, Greenbelt, Md.

Claustre, H., Morel, A., Hooker, S., Babin, M., Antoine, D., Oubelkheir, K., Bricaud, A., Leblanc, K., Queguiner, B., and Maritorena, S. (2002). Is desert dust making oligotrophic waters greener? *Geophysical Research Letters*, 29(10).

Cococcioni, M., Corsini, G., Lazzerini, B., and Marcelloni, F. (2004). Approaching the Ocean Color problem using fuzzy rules. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, 34(3):1360–73.

Cooper, J. (2001). Geomorphological variability among microtidal estuaries from the wave-dominated South African coast. *Geomorphology*, 40(1):99–122.

Cota, G. F., Harrison, W. G., Platt, T., Sathyendranath, S., and Stuart, V. (2003). Bio-optical properties of the Labrador Sea. *Journal of Geophysical Research: Oceans (1978–2012)*, 108(C7).

Cox, C. and Munk, W. (1954a). Measurement of the roughness of the sea surface from photographs of the sun's glitter. *JOSA*, 44(11):838–850.

Cox, C. and Munk, W. (1954b). Statistics of the sea surface derived from sun glitter. *Journal of Marine Research*, 13(2):198–227.

Craig, S. E., Jones, C. T., Li, W. K., Lazin, G., Horne, E., Caverhill, C., and Cullen, J. J. (2012). Deriving optical metrics of coastal phytoplankton biomass from ocean colour. *Remote Sensing of Environment*, 119:72–83.

Crichton, M., Hutchings, L., Lamont, T., and Jarre, A. (2013). From physics to phytoplankton: prediction of dominant cell size in St Helena Bay in the Southern Benguela. *Journal of plankton research*, page fbt012.

Cristina, S. C. V., Moore, G. F., Goela, P. R. F. C., Icely, J. D., and Newton, A. (2014). In situ validation of MERIS marine reflectance off the southwest Iberian Peninsula: assessment of vicarious adjustment and corrections for near-land adjacency. *International Journal of Remote Sensing*, 35(6):2347–2377.

Cyrus, D. and Blaber, S. (1987). The influence of turbidity on juvenile marine fishes in estuaries. Part 1. Field studies at Lake St. Lucia on the southeastern coast of Africa. *Journal of Experimental Marine Biology and Ecology*, 109(1):53–70.

Dall'Olmo, G. and Gitelson, A. A. (2005). Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Applied Optics*, 44(3):412–422.

Dall'Olmo, G. and Gitelson, A. A. (2006). Effect of bio-optical parameter variability and uncertainties in reflectance measurements on the remote estimation of chlorophyll-a concentration in turbid productive waters: modeling results. *Applied Optics*, 45(15):3577–3592.

Dall'Olmo, G., Gitelson, A. A., and Rundquist, D. C. (2003). Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. *Geophysical Research Letters*, 30(18).

Davis, C. O., Kavanaugh, M., Letelier, R., Bissett, W. P., and Kohler, D. (2007). Spatial and spectral resolution considerations for imaging coastal waters. In *Optical Engineering+ Applications*, pages 66800P–66800P. International Society for Optics and Photonics.

de Ruijter, W. P., van Leeuwen, P. J., and Lutjeharms, J. R. (1999). Generation and evolution of Natal Pulses: solitary meanders in the Agulhas Current. *Journal of physical oceanography*, 29(12):3043–3055.

Defoin-Platel, M. and Chami, M. (2007). How ambiguous is the inverse problem of ocean color in coastal waters? *Journal of Geophysical Research*, 112(C3):1–16.

Demarcq, H., Barlow, R., and Hutchings, L. (2007). Application of a chlorophyll index derived from satellite data to investigate the variability of phytoplankton in the Benguela ecosystem. *African Journal of Marine Science*, 29(2):271–282.

Demarcq, H., Barlow, R. G., and Shillington, F. a. (2003). Climatology and Variability of Sea Surface Temperature and Surface Chlorophyll in the Benguela and Agulhas Ecosystems As Observed by Satellite Imagery. *African Journal of Marine Science*, 25(1):363–372.

Dierssen, H. M. (2010). Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40):17073–8.

Dierssen, H. M., Kudela, R. M., Ryan, J. P., and Zimmerman, R. C. (2006). Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments. *Limnology and Oceanography*, 51(6):2646–2659.

Dierssen, H. M., Zimmerman, R. C., Leathers, R. A., Downes, T. V., and Davis, C. O. (2003). Ocean color remote sensing of seagrass and bathymetry in the Bahamas Banks by high-resolution airborne imagery. *Limnology and Oceanography*, 48(1part2):444–455.

Doerffer, R. and Schiller, H. (2007). The MERIS Case 2 water algorithm. *International Journal of Remote Sensing*, 28(3-4):517–535.

Doerffer, R. and Schiller, H. (2008). MERIS regional coastal and lake case 2 water project atmospheric correction ATBD. *GKSS research center*, 21502.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., et al. (2012). The global monitoring for environment and security (GMES) sentinel-3 mission. *Remote Sensing of Environment*, 120:37–57.

d'Ortenzio, F. and Ribera d'Alcalà, M. (2009). On the trophic regimes of the Mediterranean Sea: a satellite analysis. *Biogeosciences*, 6(2):139–148.

D'Sa, E., Hu, C., Muller-Karger, F., and Carder, K. (2002). Estimation of colored dissolved organic matter and salinity fields in case 2 waters using SeaWiFS: Examples from Florida Bay and Florida Shelf. *Journal of Earth System Science*, 111(3):197–207.

D'Sa, E. J., Miller, R. L., and Del Castillo, C. (2006). Bio-optical properties and ocean color algorithms for coastal waters influenced by the Mississippi River during a cold front. *Applied Optics*, 45(28):7410–7428.

DWA (2016). *Data from the Hydrological Information System and Peak Flows (monthly and hydrological year)*. https://www.dwa.gov.za/hydrology/HyDataSets.aspx?Station=V5H002 [2016-06-06]. Accessed: 2016-07-27.

Dwarakish, G. and Nithyapriya, B. (2016). Application Of Soft Computing Techniques In Coastal Study âĂŞ A Review. *Journal of Ocean Engineering and Science*, pages –.

Dwivedi, R., Rafeeq, M., Smitha, B. R., Padmakumar, K. B., Thomas, L. C., Sanjeevan, V. N., Prakash, P., and Raman, M. (2015). Species identification of mixed algal bloom in

the Northern Arabian Sea using remote sensing techniques. *Environmental Monitoring and Assessment*, 187.

El-habashi, A. and Ahmed, S. (2015). Neural network algorithms for retrieval of harmful algal blooms in the west Florida shelf from VIIRS satellite observations and comparisons with other techniques, without the need for a fluorescence channel. In *SPIE Remote Sensing*, pages 96380B–96380B. International Society for Optics and Photonics.

Evers-King, H. (2014). *Phytoplankton community structure determined through remote sensing and in situ optical measurements*. PhD thesis, University of Cape Town.

Evers-King, H., Bernard, S., Lain, L. R., and Probyn, T. A. (2014). Sensitivity in reflectance attributed to phytoplankton cell size: forward and inverse modelling approaches. *Optics express*, 22(10):11536–11551.

Fawcett, A., Pitcher, G., and Shillington, F. (2008). Nearshore currents on the southern Namaqua shelf of the Benguela upwelling system. *Continental Shelf Research*, 28(8):1026–1039.

Fawcett, R., Pitcher, G., Bernard, S., Cembella, A., Kudela, R., et al. (2007). Contrasting wind patterns and toxigenic phytoplankton in the southern Benguela upwelling system. *Marine Ecology-Progress Series*, 348:19–31.

Feldman, G. and McClain, C. (2015). *Chlorophyll a (chlor_a)*. http://oceancolor.gsfc.nasa.gov/cms/atbd/chlor_a [2015-05-04]. Accessed: 2015-11-24.

Feng, H., Campbell, J., Dowell, M., and Moore, T. (2005). Modeling spectral reflectance of optically complex waters using bio-optical measurements from Tokyo Bay. *Remote Sensing of Environment*, 99:232–243.

Fichot, C. G., Sathyendranath, S., and Miller, W. L. (2008). SeaUV and SeaUV C: Algorithms for the retrieval of UV/Visible diffuse attenuation coefficients from ocean color. *Remote Sensing of Environment*, 112(4):1584–1602.

Firestone, E. R. and Hooker, S. B. (1998). SeaWiFS Technical Report Series. Volume 43: SeaWiFS Prelaunch Technical Report Series Final Cumulative Index. Technical Memorandum 104566, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

Fiscella, B., Giancaspro, A., Nirchio, F., Pavese, P., and Trivero, P. (2000). Oil spill detection using marine SAR images. *International Journal of Remote Sensing*, 21(18):3561–3566.

Flemming, B. (1981). Factors controlling shelf sediment dispersal along the southeast African continental margin. *Developments in Sedimentology*, 32:259–277.

Flink, P., Lindell, T., and Ostlund, C. (2001). Statistical analysis of hyperspectral data from two Swedish lakes. *The Science of the total environment*, 268(1-3):155–69.

Franz, B. A., Kwiatowska, E. J., Meister, G., and McClain, C. R. (2008). Moderate Resolution Imaging Spectroradiometer on Terra: limitations for ocean color applications. *Journal of Applied Remote Sensing*, 2(1):023525–023525.

Franz, B. A. and Werdell, P. J. (2010). A generalized framework for modeling of inherent optical properties in ocean remote sensing applications. *Proceedings of Ocean Optics, Anchorage, Alaska*, 27.

Garver, S. A. and Siegel, D. A. (1997). Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: 1. Time series from the Sargasso Sea. *Journal of Geophysical Research: Oceans (1978–2012)*, 102(C8):18607–18625.

Garzoli, S. and Gordon, A. (1996). Origins and variability of the Benguela Current. *Journal of Geophysical Research: Oceans*, 101(C1):897–906.

Giardino, C., Bresciani, M., Villa, P., and Martinelli, A. (2010). Application of remote sensing in water resource management: the case study of Lake Trasimeno, Italy. *Water resources management*, 24(14):3885–3899.

Gilerson, A., Zhou, J., Hlaing, S., Ioannou, I., Gross, B., Moshary, F., and Ahmed, S. (2008). Fluorescence component in the reflectance spectra from coastal waters. II. Performance of retrieval algorithms. *Optics express*, 16(4):2446–2460.

Gilerson, A. a., Gitelson, A. a., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., and Ahmed, S. a. (2010). Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. *Optics express*, 18:24109–24125.

Gitelson, A. (1992). The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. *International Journal of Remote Sensing*, 13(17):3367–3373.

Gitelson, A. A., Dall'Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., Gurlin, D., and Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, 112(9):3582–3593.

Gitelson, A. a., Gao, B.-C., Li, R.-R., Berdnikov, S., and Saprygin, V. (2011a). Estimation of chlorophyll-a concentration in productive turbid waters using a Hyperspectral Imager for the Coastal Ocean - the Azov Sea case study. *Environmental Research Letters*, 6(2):024023.

Gitelson, A. A., Gurlin, D., Moses, W., and Yacobi, Y. (2011b). Remote estimation of chlorophyll-a concentration in inland, estuarine, and coastal waters. In *Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications*. CRC Press Taylor and Francis Group.

González Vilas, L., Spyrakos, E., and Torres Palenzuela, J. M. (2011). Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician rias (NW Spain). *Remote Sensing of Environment*, 115(2):524–535.

Gordon, H. and Morel, A. (1983). *Remote assessment of ocean color for satellite visible imagery. A review*. New York: Springer.

Gordon, H. R. and Clark, D. K. (1980). Remote sensing optical properties of a stratified ocean: an improved interpretation. *Applied Optics*, 19(20):3428–3430.

Goschen, W. and Schumann, E. (1988). Ocean current and temperature structures in Algoa Bay and beyond in November 1986. *South African Journal of Marine Science*, 7(1):101–116.

Goschen, W., Schumann, E., Bernard, K., Bailey, S., and Deyzel, S. (2012). Upwelling and ocean structures off Algoa Bay and the south-east coast of South Africa. *African Journal of Marine Science*, 34(4):525–536.

Gould, R. and Arnone, R. (2003). Optical water mass classification for ocean color imagery. In *Second International Conference, Current Problems in Optics Of Natural Waters. St. Petersburg, Russia*.

Govender, M., Chetty, K., and Bulcock, H. (2007). A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA*, 33(2).

Gower, J., Doerffer, R., and Borstad, G. (1999). Interpretation of the 685nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by meris. *International Journal of Remote Sensing*, 20(9):1771–1786.

Gower, J. and King, S. (2007). Validation of chlorophyll fluorescence derived from MERIS on the west coast of Canada. *International Journal of Remote Sensing*, 28(3-4):625–635.

Gower, J., King, S., Borstad, G., and Brown, L. (2005). Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *International Journal of Remote Sensing*, 26(9):2005–2012.

Goyens, C., Jamet, C., and Schroeder, T. (2013). Evaluation of four atmospheric correction algorithms for MODIS-Aqua images over contrasted coastal waters. *Remote Sensing of Environment*, 131:63–75.

Grant, M., Jackson, T., Chuprin, A., Sathyendranath, S., Zühlke, M., Storm, T., Boettcher, M., and Fomferra, N. (2015). Ocean Colour Climate Change Initiative (OC-CCI) - Phase Two Product User Guide. Issue 2.0.5.

Gregg, W. W. and Carder, K. L. (1990). A simple spectral solar irradiance model for cloudless maritime atmospheres. *Limnology and oceanography*, 35(8):1657–1675.

Gregg, W. W. and Casey, N. W. (2007). Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. *Remote Sensing of Environment*, 111(1):25–35.

Guanter, L., Ruiz-Verdú, A., Odermatt, D., Giardino, C., Simis, S., Estellés, V., Heege, T., Domínguez-gómez, J. A., and Moreno, J. (2010). Atmospheric correction of ENVISAT/MERIS data over inland waters: Validation for European lakes. *Remote Sensing of Environment*, 114(3):467–480.

Gurlin, D., Gitelson, A. A., and Moses, W. J. (2011). Remote estimation of chl-a concentration in turbid productive waters - Return to a simple two-band NIR-red model? *Remote Sensing of Environment*, 115(12):3479–3490.

Harris, S., Cyrus, D., and Beckley, L. (2001). Horizontal trends in larval fish diversity and abundance along an ocean-estuarine gradient on the northern KwaZulu-Natal coast, South Africa. *Estuarine, Coastal and Shelf Science*, 53(2):221–235.

Harrison, A. and Coombes, C. (1988). An opaque cloud cover model of sky short wavelength radiance. *Solar Energy*, 41(4):387–392.

Harrison, M. (1984). A generalized classification of South African summer rain-bearing synoptic systems. *Journal of Climatology*, 4(5):547–560.

Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley & Sons, Inc.

Hoffbeck, J. P. and Landgrebe, D. A. (1996). Covariance matrix estimation and classification with limited training data. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):763–767.

Hoge, F. E., Wright, C. W., Lyon, P. E., Swift, R. N., and Yungel, J. K. (1999). Satellite retrieval of inherent optical properties by inversion of an oceanic radiance model: a preliminary algorithm. *Applied Optics*, 38(3):495–504.

Hooker, S. B., Firestone, E. R., Acker, J. G., Campbell, J. W., Blaisdell, J. M., and Darzi, M. (1995). SeaWiFS technical report series. Volume 32: Level-3 SeaWiFS data products. Spatial and temporal binning algorithms. Technical Memorandum 104566, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

Hu, C., Lee, Z., and Franz, B. (2012). Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research: Oceans (1978–2012)*, 117(C1).

Hu, C., Lee, Z., Ma, R., Yu, K., Li, D., and Shang, S. (2010). Moderate resolution imaging spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. *Journal of Geophysical Research: Oceans (1978–2012)*, 115(C4).

Hutchings, L., Beckley, L., Griffiths, M., Roberts, M., Sundby, S., and Van der Lingen, C. (2002). Spawning on the edge: spawning grounds and nursery areas around the southern African coastline. *Marine and Freshwater Research*, 53(2):307–318.

Hutchings, L., Morris, T., Van Der Lingen, C., Lamberth, S., Connell, A., Taljaard, S., and van Niekerk, L. (2010). Ecosystem considerations of the KwaZulu-Natal sardine run. *African Journal of Marine Science*, 32(2):413–421.

Hutchings, L., Van der Lingen, C., Shannon, L., Crawford, R., Verheye, H., Bartholomae, C., Van der Plas, A., Louw, D., Kreiner, A., Ostrowski, M., et al. (2009). The Benguela Current: An ecosystem of four components. *Progress in Oceanography*, 83(1):15–32.

Iglesias-Rodríguez, M. D., Brown, C. W., Doney, S. C., Kleypas, J., Kolber, D., Kolber, Z., Hayes, P. K., and Falkowski, P. G. (2002). Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids. *Global Biogeochemical Cycles*, 16(4):47–1.

Iida, T., Mizobata, K., and Saitoh, S.-I. (2012). Interannual variability of coccolithophore *Emiliania huxleyi* blooms in response to changes in water column stability in the eastern Bering Sea. *Continental Shelf Research*, 34:7–17.

Iida, T., Saitoh, S., Miyamura, T., Toratani, M., Fukushima, H., and Shiga, N. (2002). Temporal and spatial variability of coccolithophore blooms in the eastern bering sea, 1998-2001. *Progress in Oceanography*, 55(1):165–175.

IOCCG (2000). Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters. In Sathyendranath, S., editor, *Reports of the International Ocean-Colour Coordinating Group, No. 3*, IOCCG, Dartmouth, Canada.

IOCCG (2006). Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications. In Lee, Z., editor, *Reports of the International Ocean-Colour Coordinating Group, No. 5*, IOCCG, Dartmouth, Canada.

IOCCG (2008). Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology. In Platt, T., Hoepffner, N., Stuart, V., and Brown, C., editors, *Reports of the International Ocean-Colour Coordinating Group, No. 7*, IOCCG, Dartmouth, Canada.

IOCCG (2009a). Partition of the Ocean into Ecological Provinces: Role of Ocean-Colour Radiometry. In Dowell, M. and Platt, T., editors, *Reports of the International Ocean-Colour Coordinating Group, No. 9*, IOCCG, Dartmouth, Canada.

IOCCG (2009b). Remote Sensing in Fisheries and Aquaculture. In Forget, M.-H., Stuart, V., and Platt, T., editors, *Reports of the International Ocean-Colour Coordinating Group, No. 8*, IOCCG, Dartmouth, Canada.

IOCCG (2010). Atmospheric Correction for Remotely-Sensed Ocean-Colour Products. In Wang, M., editor, *Reports of the International Ocean-Colour Coordinating Group, No. 10*, IOCCG, Dartmouth, Canada.

IOCCG (2012). Mission Requirements for Future Ocean-Colour Sensors. In McClain, C. R. and Meister, G., editors, *Reports of the International Ocean-Colour Coordinating Group, No. 13*, IOCCG, Dartmouth, Canada.

IOCCG (2014). Phytoplankton Functional Types from Space. In Sathyendranath, S., editor, *Reports of the International Ocean-Colour Coordinating Group, No. 15*, IOCCG, Dartmouth, Canada.

Jackson, J. M., Rainville, L., Roberts, M. J., McQuaid, C. D., and Lutjeharms, J. R. (2012). Mesoscale bio-physical interactions between the Agulhas Current and the Agulhas Bank, South Africa. *Continental Shelf Research*, 49:10–24.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323.

Jaquemet, S., Le Corre, M., and Quartly, G. (2007). Ocean control of the breeding regime of the sooty tern in the southwest Indian Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 54(1):130–142.

Jarre, A., Hutchings, L., Kirkman, S. P., Kreiner, A., Tchipalanga, P., Kainge, P., Uanivi, U., van der Plas, A. K., Blamey, L. K., Coetzee, J. C., et al. (2015). Synthesis: climate effects on biodiversity, abundance and distribution of marine organisms in the Benguela. *Fisheries Oceanography*, 24(S1):122–149.

Jerlov, N. G. (1951). Optical Studies of Ocean Water. *Rep. Swedish Deep-Sea Expedition*, 3:1–59.

Jerlov, N. G. (1976). *Marine optics*. Elsevier.

Jolivet, D., Ramon, D., Deschamps, P.-Y., Steinmetz, F., Fougnie, B., and Henry, P. (2007). How the ocean color product quality is limited by atmospheric correction. In Lacoste, H. and Ouwehand, L., editors, *Proceedings of the Envisat Symposium 2007*. ESA Communication Production Office, ESTEC, Noordwijk, The Netherlands.

Kahru, M., Jacox, M. G., Lee, Z., Kudela, R. M., Manzano-Sarabia, M., and Mitchell, B. G. (2015). Optimized multi-satellite merger of primary production estimates in the California Current using inherent optical properties. *Journal of Marine Systems*, 147:94–102.

Karabashev, G. and Evdoshenko, M. (2016). Narrowband shortwave minima in spectra of backscattered light from the sea obtained from ocean color scanners as a remote indication of algal blooms. *Oceanologia*.

Karabashev, G., Evdoshenko, M., and Sheberstov, S. (2006). Normalized radiance spectrum as a water exchange event diagnostic. *International Journal of Remote Sensing*, 27(9):1775–1792.

Kay, S., Hedley, J. D., and Lavender, S. (2009). Sun glint correction of high and low spatial resolution images of aquatic scenes: A review of methods for visible and near-infrared wavelengths. *Remote Sensing*, 1(4):697–730.

Kirk, J. (1980). Spectral adsorption properties of natural watres: contribtion of the soluble and particulate fractions to light absorption in some inland watres of south-eastern Australia. *Marine and Freshwater Research*, 31(3):287–296.

Kirkman, S., Blamey, L., Lamont, T., Field, J., Bianchi, G., Huggett, J., Hutchings, L., Jackson-Veitch, J., Jarre, A., Lett, C., et al. (2016). Spatial characterisation of the Benguela ecosystem for ecosystem-based management. *African Journal of Marine Science*, pages 1–16.

Knap, A., Michaels, A., Close, A., Ducklow, H., and Dickson, A. (1996). Protocols for the joint global ocean flux study (JGOFS) core measurements. *JGOFS, Reprint of the IOC Manuals and Guides No. 29, UNESCO 1994*, 19.

Kohonen, T. (1984). *Self-organization and associative memory*, volume 2. (p. 312). Springer-Verlag, Berlin.

Koponen, S., Attila, J., Pulliainen, J., Kallio, K., Pyhälahti, T., Lindfors, A., Rasmus, K., and Hallikainen, M. (2007). A case study of airborne and satellite remote sensing of a spring bloom event in the Gulf of Finland. *Continental Shelf Research*, 27(2):228–244.

Krawczyk, H., Ebert, K., and Neumann, A. (2004). Algae Bloom detection in the Baltic Sea with MERIS data. In *Proc. MERIS User Workshop*.

Krug, M., Cipollini, P., and Dufois, F. (2014a). Observing the Agulhas Current with sea surface temperature and altimetry data: challenges and perspectives. In *Remote Sensing of the African Seas*, pages 233–249. Springer.

Krug, M., Tournadre, J., and Dufois, F. (2014b). Interactions between the Agulhas Current and the eastern margin of the Agulhas Bank. *Continental Shelf Research*, 81:67–79.

Kuchinke, C. P., Gordon, H. R., and Franz, B. a. (2009). Spectral optimization for constituent retrieval in Case 2 waters I: Implementation and performance. *Remote Sensing of Environment*, 113(3):571–587.

Kutser, T. (2004). Quantitative Detection of Chlorophyll in Cyanobacterial Blooms by Satellite Remote Sensing. *Limnology and Oceanography*, 49(6):2179–2189.

Kutser, T., Metsamaa, L., Strömbeck, N., and Vahtmäe, E. (2006). Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine, Coastal and Shelf Science*, 67(1-2):303–312.

Kwiatkowska, E. J., Franz, B. A., Meister, G., McClain, C. R., and Xiong, X. (2008). Cross calibration of ocean-color bands from Moderate Resolution Imaging Spectroradiometer on Terra platform. *Applied Optics*, 47(36):6796–6810.

Lamont, T. and Barlow, R. (2015). Environmental influence on phytoplankton production during summer on the KwaZulu-Natal shelf of the Agulhas ecosystem. *African Journal of Marine Science*, 37(4):485–501.

Lamont, T., Barlow, R., and Kyewalyanga, M. (2014). Physical drivers of phytoplankton production in the southern Benguela upwelling system. *Deep Sea Research Part I: Oceanographic Research Papers*, 90:1–16.

Largier, J., Chapman, P., Peterson, W., and Swart, V. (1992). The western Agulhas Bank: circulation, stratification and ecology. *South African Journal of Marine Science*, 12(1):319–339.

Lavender, S. and Kay, S. (2010). OLCI Level 2 ATBD Glint correction. Issue 2.0.

Le, C., Li, Y., Zha, Y., Sun, D., Huang, C., and Zhang, H. (2011). Remote estimation of chlorophyll a in optically complex waters based on optical classification. *Remote Sensing of Environment*, 115(2):725–737.

Lee, Z., Arnone, R., Hu, C., Werdell, P. J., and Lubac, B. (2010). Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Applied Optics*, 49(3):369–381.

Lee, Z., Carder, K. L., and Arnone, R. a. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Applied Optics*, 41(27):5755–72.

Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G., and Patch, J. S. (1999). Hyperspectral remote sensing for shallow waters. 2. Deriving bottom depths and water properties by optimization. *Applied Optics*, 38(18):3831–43.

Lee-Thorp, A., Rouault, M., and Lutjeharms, J. (1998). Cumulus cloud formation above the Agulhas Current. *South African journal of science*, 94(7):351–354.

Lerebourg, C. and Bruniquel, V. (2011). MERIS 3rd data reprocessing Software and ADF updates. In *European Space Agency Report. Ref, A879. NT. 008. ACRI-ST.*

Letelier, R. M. and Abbott, M. R. (1996). An analysis of chlorophyll fluorescence algorithms for the Moderate Resolution Imaging Spectrometer (MODIS). *Remote Sensing of Environment*, 58(2):215–223.

Lett, C., Roy, C., Levasseur, A., Van Der Lingen, C. D., and Mullon, C. (2006). Simulation and quantification of enrichment and retention processes in the southern Benguela upwelling ecosystem. *Fisheries Oceanography*, 15(5):363–372.

Li, Y., Wang, Q., Wu, C., Zhao, S., Xu, X., Wang, Y., and Huang, C. (2012). Estimation of chlorophyll a concentration using NIR/red bands of MERIS and classification procedure in inland turbid water. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(3):988–997.

Lihan, T., Saitoh, S.-I., Iida, T., Hirawake, T., and Iida, K. (2008). Satellite-measured temporal and spatial variability of the Tokachi River plume. *Estuarine, Coastal and Shelf Science*, 78(2):237–249.

Liu, D. and Wang, Y. (2013). Trends of satellite derived chlorophyll-a (1997-2011) in the Bohai and Yellow Seas, China: Effects of bathymetry on seasonal and inter-annual patterns. *Progress in Oceanography*, 116:154–166.

Longhurst, A. (1998). *Ecological Geography of the Sea*. Academic Press, San Diego.

Louchard, E. M., Reid, R. P., Stephens, F. C., Davis, C. O., Leathers, R. a., and Downes, T. V. (2003). Optical remote sensing of benthic habitats and bathymetry in coastal environments at Lee Stocking Island, Bahamas: A comparative spectral classification approach. *Limnology and Oceanography*, 48(1_part_2):511–521.

Lubac, B. and Loisel, H. (2007). Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. *Remote Sensing of Environment*, 110(1):45–58.

Lucas, A. J., Pitcher, G. C., Probyn, T. A., and Kudela, R. M. (2014). The influence of diurnal winds on phytoplankton dynamics in a coastal upwelling system off southwestern Africa. *Deep Sea Research Part II: Topical Studies in Oceanography*, 101:50–62.

Lutjeharms, J. (2007). Three decades of research on the greater Agulhas Current. *Ocean Science*, 3(1):129–147.

Lutjeharms, J., Cooper, J., and Roberts, M. (2000a). Upwelling at the inshore edge of the Agulhas Current. *Continental Shelf Research*, 20(7):737–761.

Lutjeharms, J., Mey, R., and Hunter, I. (1986). Cloud lines over the Agulhas Current. *South African Journal of Science*, 82(11):635–640.

Lutjeharms, J. and Roberts, H. (1988). The Natal pulse: An extreme transient on the Agulhas Current. *Journal of Geophysical Research*, 93(C1):631–645.

Lutjeharms, J. and Rouault, M. (2000). Observations of cloud formation above Agulhas current intrusions in the southeast Atlantic. *South African Journal of Science*, 96(11/12):577–579.

Lutjeharms, J., Valentine, H., and Van Ballegooyen, R. (2000b). The hydrography and water masses of the Natal Bight, South Africa. *Continental Shelf Research*, 20(14):1907–1939.

Lutjeharms, J. R. (2006). *The Agulhas Current*. Springer-Verlag, Berlin, Heidelberg.

Machu, E., Ferret, B., and Garçon, V. (1999). Phytoplankton pigment distribution from SeaWiFS data in the subtropical convergence zone south of Africa: A wavelet analysis. *Geophysical Research Letters*, 26(10):1469.

Machu, E. and Garçon, V. (2001). Phytoplankton seasonal distribution from SeaWiFS data in the Agulhas Current system. *Journal of marine research*, 59(5):795–812.

Magnuson, A., Harding, L. W., Mallonee, M. E., and Adolf, J. E. (2004). Bio-optical model for Chesapeake Bay and the middle Atlantic bight. *Estuarine, Coastal and Shelf Science*, 61(3):403–424.

Mao, J. and Jain, A. K. (1996). A self-organizing network for hyperellipsoidal clustering (HEC). *IEEE Transactions on Neural Networks*, 7(1):16–29.

Mao, Z., Chen, J., Hao, Z., Pan, D., Tao, B., and Zhu, Q. (2013). A new approach to estimate the aerosol scattering ratios for the atmospheric correction of satellite remote sensing data in coastal regions. *Remote Sensing of Environment*, 132:186–194.

Margalef, R. (1962). Succession in marine populations. *Advancing Frontiers of Plant Sciences*, 2:137–188.

Maritorena, S., Siegel, D. A., and Peterson, A. R. (2002). Optimization of a semianalytical ocean color model for global-scale applications. *Applied Optics*, 41(15):2705–14.

Marrari, M., Hu, C., and Daly, K. (2006). Validation of SeaWiFS chlorophyll a concentrations in the Southern Ocean: A revisit. *Remote Sensing of Environment*, 105(4):367–375.

Matsuoka, A., Hooker, S. B., Bricaud, A., Gentili, B., and Babin, M. (2013). Estimating absorption coefficients of colored dissolved organic matter (CDOM) using a semi-analytical algorithm for southern Beaufort Sea waters: application to deriving concentrations of dissolved organic carbon from space. *Biogeosciences*, 10(2):917–927.

Matsushita, B., Yang, W., Yu, G., Oyama, Y., Yoshimura, K., and Fukushima, T. (2015). A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in Asian inland waters. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102:28–37.

Matthews, M. W. (2014). Eutrophication and cyanobacterial blooms in South African inland waters: 10years of MERIS observations. *Remote Sensing of Environment*, 155:161–177.

Matthews, M. W., Bernard, S., and Robertson, L. (2012). An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sensing of Environment*, 124:637–652.

Matthews, M. W., Bernard, S., and Winter, K. (2010). Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Remote Sensing of Environment*, 114(9):2070–2087.

Mayot, N., D'Ortenzio, F., d'Alcalà, M. R., Lavigne, H., and Claustre, H. (2015). Interannual variability of the Mediterranean trophic regimes from ocean color satellites. *Biogeosciences Discussions*, 12:14941–14980.

McClain, C. R. (2009). A decade of satellite ocean color observations*. *Annual Review of Marine Science*, 1:19–42.

McClain, C. R., Meister, G., and Monosmith, B. (2014). Chapter 2.1 - Satellite Ocean Color Sensor Design Concepts and Performance Requirements. In Giuseppe Zibordi, C. J. D. and Parr, A. C., editors, *Optical Radiometry for Ocean Climate Measurements*, volume 47 of *Experimental Methods in the Physical Sciences*, pages 73 – 119. Academic Press.

McKee, D., Cunningham, A., and Dudek, A. (2007a). Optical water type discrimination and tuning remote sensing band-ratio algorithms: Application to retrieval of chlorophyll and $K_d(490)$ in the Irish and Celtic Seas. *Estuarine, Coastal and Shelf Science*, 73(3):827–834.

McKee, D., Cunningham, A., Wright, D., and Hay, L. (2007b). Potential impacts of nonalgal materials on water-leaving Sun induced chlorophyll fluorescence signals in coastal waters. *Applied Optics*, 46(31):7720–7729.

McKinna, L. I., Furnas, M. J., and Ridd, P. V. (2011). A simple, binary classification algorithm for the detection of Trichodesmium spp. within the Great Barrier Reef using MODIS imagery. *Limnology and Oceanography: Methods*, 9(2):50–66.

Meister, G., Eplee, R. E., and Franz, B. A. (2014). Corrections to MODIS Terra calibration and polarization trending derived from ocean color products. In *SPIE Optical Engineering+ Applications*, pages 92180V–92180V. International Society for Optics and Photonics.

Meister, G. and Franz, B. A. (2011). Adjustments to the MODIS Terra radiometric calibration and polarization sensitivity in the 2010 reprocessing. In *SPIE Optical Engineering+ Applications*, pages 815308–815308. International Society for Optics and Photonics.

Mélin, F., Sclep, G., Jackson, T., and Sathyendranath, S. (2016). Uncertainty estimates of remote sensing reflectance derived from comparison of ocean color satellite data sets. *Remote Sensing of Environment*, 177:107–124.

Mélin, F. and Vantrepotte, V. (2015). How optically diverse is the coastal ocean? *Remote Sensing of Environment*, 160:235–251.

Mélin, F., Vantrepotte, V., Clerici, M., D'Alimonte, D., Zibordi, G., Berthon, J.-F., and Canuti, E. (2011). Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the adriatic sea. *Progress in Oceanography*, 91(3):229–244.

Merico, a. (2003). Analysis of satellite imagery for *Emiliania huxleyi* blooms in the Bering Sea before 1997. *Geophysical Research Letters*, 30(6):1337.

Meyer, A., Lutjeharms, J., and de Villiers, S. (2002). The nutrient characteristics of the Natal Bight, South Africa. *Journal of Marine Systems*, 35(1-2):11–37.

Mitchell, B., Brody, E. A., Yeh, E.-n., Mcclain, C., Comiso, J., and Maynard, N. G. (1991). Meridional zonation of the Barents Sea ecosystem inferred from satellite remote sensing and in situ bio-optical observations. *Polar research*, 10(1):147–162.

Mitchell, B. G. and Kahru, M. (2009). Bio-optical algorithms for ADEOS-2 GLI. *Journal of The Remote Sensing Society of Japan*, 29(1):80–85.

Mitchell-Innes, B., Richardson, A., and Painting, S. (1999). Seasonal changes in phytoplankton biomass on the western Agulhas Bank, South Africa. *South African Journal of Marine Science*, 21(1):217–233.

Mobley, C. (2010). *Overview of Optical Oceanography: Apparent Optical Properties*. http://www.oceanopticsbook.info/view/overview_of_optical_oceanography/apparent_optical_properties [2010-02-03]. Accessed: 2016-08-02.

Mobley, C., Stramski, D., Bissett, W., and Boss, E. (2004). Optical modeling of ocean waters: Is the Case 1-Case 2 classification still useful? *Oceanography*, 17(2):60–67.

Mobley, C. and Sundman, L. (2008). Hydrolight 5 ecolight 5 technical documentation. *Sequoia Scientific, Incorporated, Bellevue, WA*, 98005:95.

Mobley, C. D. (1994). *Light and water: Radiative transfer in natural waters*. Academic press.

Mobley, C. D. (2011). Fast light calculations for ocean ecosystem and inverse models. *Optics express*, 19(20):18927–18944.

Mobley, C. D., Gentili, B., Gordon, H. R., Jin, Z., Kattawar, G. W., Morel, A., Reinersman, P., Stamnes, K., and Stavn, R. H. (1993). Comparison of numerical models for computing underwater light fields. *Applied Optics*, 32(36):7484–7504.

Mobley, C. D., Sundman, L. K., Davis, C. O., Bowles, J. H., Downes, T. V., Leathers, R. a., Montes, M. J., Bissett, W. P., Kohler, D. D. R., Reid, R. P., Louchard, E. M., and Gleason, A. (2005). Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables. *Applied Optics*, 44(17):3576–92.

Monteiro, P. M. and van der Plas, A. K. (2006). Low oxygen water (LOW) variability in the Benguela system: Key processes and forcing scales relevant to forecasting. *Large Marine Ecosystems*, 14:71–90.

Moore, G. and Lavender, S. (2011). Case IIS bright pixel atmospheric correction. *MERIS ATBD*, 2.

Moore, T. S., Campbell, J. W., and Dowell, M. D. (2009). A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sensing of Environment*, 113(11):2424–2430.

Moore, T. S., Campbell, J. W., and Feng, H. (2001). A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8):1764–1776.

Moore, T. S., Campbell, J. W., and Feng, H. (2015). Characterizing the uncertainties in spectral remote sensing reflectance for SeaWiFS and MODIS-Aqua based on global in situ matchup data sets. *Remote Sensing of Environment*, 159:14–27.

Moore, T. S., Dowell, M. D., Bradt, S., and Verdu, A. R. (2014). An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. *Remote Sensing of Environment*, 143:97–111.

Moore, T. S., Dowell, M. D., and Franz, B. A. (2012). Detection of coccolithophore blooms in ocean color satellite imagery: A generalized approach for use with multiple sensors. *Remote Sensing of Environment*, 117:249–263.

Morel, A. (1980). In-water and remote measurements of ocean color. *Boundary-Layer Meteorology*, 18(2):177–201.

Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters). *Journal of Geophysical Research: Oceans (1978–2012)*, 93(C9):10749–10768.

Morel, A. and Antoine, D. (2011). MERIS ATBD 2.9. Pigment index retrieval in Case 1 waters. Issue 4, Revision 3.

Morel, A., Antoine, D., and Gentili, B. (2002). Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function. *Applied Optics*, 41(30):6289–6306.

Morel, A. and Gentili, B. (2009). The dissolved yellow substance and the shades of blue in the Mediterranean Sea. *Biogeosciences*, 6(11).

Morel, A. and Prieur, L. (1977). Analysis of variations in ocean color. *Limnology and oceanography*, 22(4):709–722.

Moses, W. J., Ackleson, S. G., Hair, J. W., Hostetler, C. A., and Miller, W. D. (2016). Spatial scales of optical variability in the coastal ocean: Implications for remote sensing and in situ sampling. *Journal of Geophysical Research: Oceans*, 121:4194–4208.

Moses, W. J., Gitelson, A., Berdnikov, S., and Povazhnyy, V. (2009a). Satellite Estimation of Chlorophyll-Concentration Using the Red and NIR Bands of MERIS - The Azov Sea Case Study. *Geoscience and Remote Sensing Letters, IEEE*, 6(4):845–849.

Moses, W. J., Gitelson, A. A., Berdnikov, S., and Povazhnyy, V. (2009b). Estimation of chlorophyll- a concentration in case II waters using MODIS and MERIS data - successes and challenges. *Environmental Research Letters*, 4(4):045005.

Moses, W. J., Gitelson, A. A., Berdnikov, S., Saprygin, V., and Povazhnyi, V. (2012). Operational MERIS-based NIR-red algorithms for estimating chlorophyll-a concentrations in coastal waters - The Azov Sea case study. *Remote Sensing of Environment*, 121:118–124.

Mueller, J., Bidigare, R., Trees, C., Balch, W., Dore, J., Drapeua, D., Karl, D., Van Heukelem, L., and Perl, J. (2003). Biogeochemical and bio-optical measurements and data analysis protocols: ocean optics protocols for satellite ocean color sensor validation. Revision 5, Vol. V. *NASA/TM-2003*, pages 15–26.

Mueller, J. L. (1976). Ocean color spectra measured off the Oregon coast: characteristic vectors. *Applied Optics*, 15(2):394–402.

Mueller, J. L. and Lange, R. E. (1989). Bio-optical provinces of the Northeast Pacific Ocean: A provisional analysis. *Limnology and Oceanography*, 34(8):1572–1586.

Nelson, G. and Hutchings, L. (1983). The Benguela upwelling area. *Progress in Oceanography*, 12(3):333–356.

Niang, a. (2003). Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge. *Remote Sensing of Environment*, 86(2):257–271.

Odermatt, D., Giardino, C., and Heege, T. (2010). Chlorophyll retrieval with MERIS Case-2-Regional in perialpine lakes. *Remote Sensing of Environment*, 114(3):607–617.

Odermatt, D., Gitelson, A., Brando, V. E., and Schaepman, M. (2012). Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sensing of Environment*, 118:116–126.

O'Donoghue, S., Drapeau, L., and Peddemors, V. (2010). Broad-scale distribution patterns of sardine and their predators in relation to remotely sensed environmental conditions during the KwaZulu-Natal sardine run. *African Journal of Marine Science*, 32(2):279–291.

Olivier, J. (2002). Fog-water harvesting along the West Coast of South Africa: A feasibility study. *Water SA*, 28(4):349–360.

O'Reilly, J., Maritorena, S., O'Brien, M., Siegel, D., Toole, D., Menzies, D., Smith, R., Mueller, J., Mitchell, B., Kahru, M., et al. (2000). SeaWiFS postlaunch calibration and validation analyses, part 3. *NASA Tech. Memo*, 2206892.

O'Reilly, J. E., Maritorena, S., Mitchell, B., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., and Mcclain, C. R. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, 103(C11):24937–24953.

Otero, M. and Siegel, D. (2004). Spatial and temporal characteristics of sediment plumes and phytoplankton blooms in the Santa Barbara Channel. *Deep Sea Research Part II: Topical Studies in Oceanography*, 51(10-11):1129–1149.

Pal, N. R. and Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, 3(3):370–379.

Pan, X., Mannino, A., Russ, M. E., and Hooker, S. B. (2008). Remote sensing of the absorption coefficients and chlorophyll a concentration in the United States southern Middle Atlantic Bight from SeaWiFS and MODIS-Aqua. *Journal of Geophysical Research: Oceans (1978–2012)*, 113(C11).

Pelevin, V. and Rutkovskaya, V. (1977). On the optical classification of the ocean waters by the spectral solar light attenuation. *Oceanology*, 17(1):15–21.

Petrenko, D., Pozdnyakov, D., Johannessen, J., Counillon, F., and Sychov, V. (2013). Satellite-derived multi-year trend in primary production in the Arctic Ocean. *International Journal of Remote Sensing*, 34(11):3903–3937.

Pitcher, G., Boyd, A., Horstman, D., and Mitchell-Innes, B. (1998). Subsurface dinoflagellate populations, frontal blooms and the formation of red tide in the southern Benguela upwelling system. *Marine Ecology Progress Series*, 172:253–264.

Pitcher, G. and Calder, D. (2000). Harmful algal blooms of the southern Benguela Current: a review and appraisal of monitoring from 1989 to 1997. *South African Journal of Marine Science*, 22(1):255–271.

Pitcher, G. and Cockcroft, A. (1998). Low oxygen, rock lobster strandings and PSP. *Harmful algae news*, (17).

Pitcher, G., Figueiras, F., Hickey, B., and Moita, M. (2010). The physical oceanography of upwelling systems and the development of harmful algal blooms. *Progress in Oceanography*, 85(1):5–32.

Pitcher, G. C. and Nelson, G. (2006). Characteristics of the surface boundary layer important to the development of red tide on the southern Namaqua shelf of the Benguela upwelling system. *Limnology and Oceanography*, 51(6):2660–2674.

Pitcher, G. C. and Probyn, T. A. (2011). Anoxia in southern Benguela during the autumn of 2009 and its linkage to a bloom of the dinoflagellate *Ceratium balechii*. *Harmful Algae*, 11:23–32.

Pitcher, G. C., Probyn, T. A., du Randt, A., Lucas, A., Bernard, S., Evers-King, H., Lamont, T., Hutchings, L., et al. (2014). Dynamics of oxygen depletion in the nearshore of a coastal embayment of the southern Benguela upwelling system. *Journal of Geophysical Research: Oceans*, 119(4):2183–2200.

Pitcher, G. C. and Weeks, S. J. (2006). The variability and potential for prediction of harmful algal blooms in the southern Benguela ecosystem. *Large Marine Ecosystems*, 14:125–146.

Pope, R. M. and Fry, E. S. (1997). Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. *Applied Optics*, 36(33):8710–8723.

Preisendorfer, R. (1976). Hydrologic optics. Vol. I. Introduction. US Department of Commerce, National Oceanic and Atmospheric Administration. *Environment Research Laboratory*.

Prieur, L. and Sathyendranath, S. (1981). An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials. *Limnology and Oceanography*, 26(4):671–689.

Probyn, T., Mitchell-Innes, B., Brown, P., Hutchings, L., and Carter, R. (1994). Review of primary production and related processes on the Agulhas-Bank. *South African Journal of Science*, 90(3):166–173.

Probyn, T. A., Pitcher, G. C., Monteiro, P. M. S., Boyd, a. J., and Nelson, G. (2000). Physical processes contributing to harmful algal blooms in Saldanha Bay, South Africa. *South African Journal of Marine Science*, 22(1):285–297.

Reinart, A., Herlevi, A., Arst, H., and Sipelgas, L. (2003). Preliminary optical classification of lakes and coastal waters in Estonia and south Finland. *Journal of Sea Research*, 49(4):357–366.

Ressom, H., Turner, K., and Musavi, M. (2006). Estimation of Ocean Water Chlorophyll-a Concentration Using Fuzzy C-means Clustering and Artificial Neural Networks. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 4118–4125.

Richards, J. (1993). *Remote Sensing Digital Image Analysis*. (p. 340) Springer-Verlag, New York.

Rivas, A. L., Dogliotti, A. I., and Gagliardini, D. A. (2006). Seasonal variability in satellite-measured surface chlorophyll in the Patagonian Shelf. *Continental Shelf Research*, 26(6):703–720.

Robertson Lain, L., Bernard, S., and Evers-King, H. (2014). Biophysical modelling of phytoplankton communities from first principles using two-layered spheres: Equivalent Algal Populations (EAP) model. *Optics express*, 22(14):16745–16758.

Roesler, C. S., Perry, M. J., and Carder, K. L. (1989). Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters. *Limnology and Oceanography*, 34(8):1510–1523.

RQIS (2016). *South Africa 1:500 000 Rivers*. https://www.dwa.gov.za/iwqs/gis_data/river/rivs500k [2016-05-24]. Accessed: 2016-07-21.

Ryan, J. P., Gower, J. F., King, S. A., Bissett, W. P., Fischer, A. M., Kudela, R. M., Kolber, Z., Mazzillo, F., Rienecker, E. V., and Chavez, F. P. (2008). A coastal ocean extreme bloom incubator. *Geophysical Research Letters*, 35(12).

Ryther, J. H., Hall, J. R., Pease, A. K., Bakun, A., and Jones, M. M. (1966). Primary organic production in relation to the chemistry and hydrography of the western Indian Ocean. *Limnology and Oceanography*, 11(3):371–380.

Santer, R. (2010). ICOL+ Algorithm Theoretical Basis Document. Technical Report V1.0. Brockmann Consult.

Santer, R. P. and Vidot, J. (2004). Atmospheric correction over turbid waters in the SISCAL project: Application to SeaWiFS. In *Remote Sensing*, pages 25–33. International Society for Optics and Photonics.

Sathyendranath, S., Cota, G., Stuart, V., Maass, H., and Platt, T. (2001). Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches. *International Journal of Remote Sensing*, 22(2-3):249–273.

Sathyendranath, S., Hoge, F. E., Platt, T., and Swift, R. N. (1994). Detection of phytoplankton pigments from ocean color: improved algorithms. *Applied Optics*, 33(6):1081–1089.

Sathyendranath, S., Longhurst, A., Caverhill, C. M., and Platt, T. (1995). Regionally and seasonally differentiated primary production in the North Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers*, 42(10):1773–1802.

Sathyendranath, S., Prieur, L., and Morel, A. (1989). A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters. *International Journal of Remote Sensing*, 10(8):1373–1394.

Sauer, M. J., Roesler, C., Werdell, P., and Barnard, A. (2012). Under the hood of satellite empirical chlorophyll a algorithms: revealing the dependencies of maximum band ratio algorithms on inherent optical properties. *Optics express*, 20(19):20920–20933.

Schaeffer, B. A., Hagy, J. D., Conmy, R. N., Lehrter, J. C., and Stumpf, R. P. (2012). An approach to developing numeric water quality criteria for coastal waters using the SeaWiFS satellite data record. *Environmental Science & Technology*, 46(2):916–922.

Schonten, M. W., Ruijter, W. P., Leeuwen, P. J., and Lutjeharms, J. R. (2000). Translation, decay and splitting of Agulhas rings in the southeastern Atlantic Ocean. *Journal of Geophysical Research: Oceans*, 105(C9):21913–21925.

Schumann, E. (1987). The coastal ocean off the east coast of South Africa. *Transactions of the Royal Society of South Africa*, 46(3):215–229.

Schumann, E. (2009). The Berg Estuary: water structures and dynamics. *Transactions of the Royal Society of South Africa*, 64(2):164–180.

Schumann, E., Perrins, L., and Hunter, I. (1982). Upwelling along the south coast of the Cape Province, South Africa. *South African Journal of Science*, 78(6):238–242.

Shannon, L. (2006). A plan comes together. *Benguela: Predicting a Large Marine Ecosystem*, 14:3–10.

Shannon, L., Hennig, H., Shillington, F., Bartels, A., and Swart, D. (1991). Colour fronts in False Bay: origin, development and implications. In *False Bay - An Environmental Assessment*, volume 47, pages 447–470.

Shannon, L., Hutchings, L., Bailey, G., and Shelton, P. (1984). Spatial and temporal distribution of chlorophyll in southern African waters as deduced from ship and satellite measurements and their implications for pelagic fisheries. *South African Journal of Marine Science*, 2(1):109–130.

Shannon, L. and Nelson, G. (1996). The Benguela: large scale features and processes and system variability. In *The South Atlantic*, pages 163–210. Springer.

Shannon, L. V. (1985). The Benguela ecosystem. I: Evolution of the Benguela physical features and processes. *Oceanography and Marine Biology*, 23:105–182.

Shen, Q., Li, J., Zhang, F., Sun, X., Li, J., Li, W., and Zhang, B. (2015). Classification of Several Optically Complex Waters in China Using in Situ Remote Sensing Reflectance. *Remote Sensing*, 7(11):14731–14756.

Shi, K., Li, Y., Li, L., Lu, H., Song, K., Liu, Z., Xu, Y., and Li, Z. (2013). Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. *Science of the Total Environment*, 444:1–15.

Silió-Calzada, A., Bricaud, A., and Gentili, B. (2008). Estimates of sea surface nitrate concentrations from sea surface temperature and chlorophyll concentration in upwelling areas: A case study for the Benguela system. *Remote Sensing of Environment*, 112(6):3173–3180.

Singh, R. K. and Shanmugam, P. (2014). A novel method for estimation of aerosol radiance and its extrapolation in the atmospheric correction of satellite data over optically complex oceanic waters. *Remote Sensing of Environment*, 142:188–206.

Siswanto, E., Ishizaka, J., Tripathy, S. C., and Miyamura, K. (2013). Detection of harmful algal blooms of Karenia mikimotoi using MODIS measurements: A case study of Seto-Inland Sea, Japan. *Remote Sensing of Environment*, 129:185–196.

Smith, M. and Pitcher, G. (2015). Saldanha Bay, South Africa I: the use of ocean colour remote sensing to assess phytoplankton biomass. *African Journal of Marine Science*, 37(4):503–512.

Smith, M. E., Bernard, S., and O'Donoghue, S. (2013). The assessment of optimal MERIS ocean colour products in the shelf waters of the KwaZulu-Natal Bight, South Africa. *Remote Sensing of Environment*, 137:124–138.

Smith, R. C. and Baker, K. S. (1978). Optical classification of natural waters. *Limnology and Oceanography*, 23(2):260–267.

Smits, P., Dellepiane, S., and Schowengerdt, R. (1999). Quality assessment of image classification algorithms for land-cover mapping: a review and a proposal for a cost-based approach. *International journal of remote sensing*, 20(8):1461–1486.

Smyth, T. J., Moore, G. F., Hirata, T., and Aiken, J. (2006). Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment. *Applied Optics*, 45(31):8116–8131.

Solberg, A. H. S., Storvik, G., Solberg, R., and Volden, E. (1999). Automatic detection of oil spills in ERS SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4):1916–1924.

Steinmetz, F., Deschamps, P.-Y., and Ramon, D. (2011). Atmospheric correction in presence of sun glint: application to MERIS. *Optics express*, 19(10):9783–9800.

Sterckx, S., Knaeps, S., Kratzer, S., and Ruddick, K. (2015). SIMilarity Environment Correction (SIMEC) applied to MERIS data over inland and coastal waters. *Remote Sensing of Environment*, 157:96–110.

Stramski, D., Boss, E., Bogucki, D., and Voss, K. J. (2004). The role of seawater constituents in light backscattering in the ocean. *Progress In Oceanography*, 61(1):27–56.

Subramaniam, A., Brown, C. W., Hood, R. R., Carpenter, E. J., and Capone, D. G. (2001). Detecting Trichodesmium blooms in SeaWiFS imagery. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(1):107–121.

Szeto, M., Werdell, P. J., Moore, T. S., and Campbell, J. W. (2011). Are the world's oceans optically different? *Journal of Geophysical Research: Oceans*, 116(C7).

Thomas, A. C. and Weatherbee, R. A. (2006). Satellite-measured temporal variability of the Columbia River plume. *Remote Sensing of Environment*, 100(2):167–178.

Tilstone, G. H., Peters, S. W., van der Woerd, H. J., Eleveld, M. A., Ruddick, K., Schönfeld, W., Krasemann, H., Martinez-Vicente, V., Blondeau-Patissier, D., Röttgers, R., et al. (2012). Variability in specific-absorption properties and their use in a semi-analytical ocean colour algorithm for MERIS in North Sea and Western English Channel Coastal Waters. *Remote Sensing of Environment*, 118:320–338.

Toole, D. A. and Siegel, D. A. (2001). Modes and mechanisms of ocean color variability in the Santa Barbara Channel. *Journal of Geophysical Research*, 106(C11):26985.

Torrecilla, E., Stramski, D., Reynolds, R. a., Millán-Núñez, E., and Piera, J. (2011). Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean. *Remote Sensing of Environment*, 115(10):2578–2593.

Traykovski, M., V, L., and Sosik, H. M. (2003). Feature-based classification of optical water types in the Northwest Atlantic based on satellite ocean color data. *Journal of Geophysical Research: Oceans (1978–2012)*, 108(C5).

Trees, C. C., Kennicutt, M. C., and Brooks, J. M. (1985). Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments. *Marine Chemistry*, 17(1):1–12.

Tzortziou, M., Herman, J. R., Gallegos, C. L., Neale, P. J., Subramaniam, A., Harding, L. W., and Ahmad, Z. (2006). Bio-optics of the Chesapeake Bay from measurements and radiative transfer closure. *Estuarine, Coastal and Shelf Science*, 68(1):348–362.

Van Ballegooyen, R., Taljaard, S., Van Niekerk, L., Lamberth, S., Theron, A., and Weerts, S. (2007). *Freshwater flow dependency in South African marine ecosystems: A proposed assessment framework and initial assessment of South African marine ecosystems*. Water Research Commission.

Van der Lingen, C., Hutchings, L., Lamont, T., and Pitcher, G. (2016). Climate change, dinoflagellate blooms and sardine in the southern Benguela Current Large Marine Ecosystem. *Environmental Development*.

van Wyk, J.-A. (2015). Defining the blue economy as a South African strategic priority: toward a sustainable 10th province? *Journal of the Indian Ocean Region*, 11(2):153–169.

Vantrepotte, V., Loisel, H., Dessailly, D., and Mériaux, X. (2012). Optical classification of contrasted coastal waters. *Remote Sensing of Environment*, 123:306–323.

Veitch, J., Penven, P., and Shillington, F. (2009). The benguela: A laboratory for comparative modeling studies. *Progress in Oceanography*, 83(1):296–302.

Verheye, H. M., Lamont, T., Huggett, J. A., Kreiner, A., and Hampton, I. (2016). Plankton productivity of the benguela current large marine ecosystem (bclme). *Environmental Development*, 17:75–92.

Volpe, G., Santoleri, R., Vellucci, V., d'Alcalà, M. R., Marullo, S., and d'Ortenzio, F. (2007). The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. *Remote Sensing of Environment*, 107(4):625–638.

Walker, N. (1986). Satellite observations of the Agulhas Current and episodic upwelling south of Africa. *Deep Sea Research Part A. Oceanographic Research Papers*, 33(8):1083–1106.

Wang, M., Son, S., and Harding, L. W. (2009). Retrieval of diffuse attenuation coefficient in the Chesapeake Bay and turbid ocean regions for satellite ocean color applications. *Journal of Geophysical Research: Oceans*, 114(C10).

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244.

Washington, R. and Todd, M. (1999). Tropical-temperate links in southern African and Southwest Indian Ocean satellite-derived daily rainfall. *International Journal of Climatology*, 19(14):1601–1616.

Wasmund, N., Nausch, G., and Hansen, A. (2014). Phytoplankton succession in an isolated upwelled Benguela water body in relation to different initial nutrient conditions. *Journal of Marine Systems*, 140:163–174.

Weeks, S., Barlow, R., Roy, C., and Shillington, F. (2006). Remotely sensed variability of temperature and chlorophyll in the southern Benguela: upwelling frequency and phytoplankton response. *African Journal of Marine Science*, 28(3-4):493–509.

Weeks, S. J., Pitcher, G. C., and Bernard, S. (2004). Satellite Monitoring of the Evolution of a Coccolithophorid Bloom in the Southern Benguela Upwelling System. *Oceanography*, 17(1):83–89.

Werdell, P. J. and Bailey, S. W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, 98(1):122–140.

Windham, M. P. (1982). Cluster Validity for the Fuzzy c-Means Clustering Algorithrm. *IEEE transactions on pattern analysis and machine intelligence*, 4(4):357–63.

Xie, X. L. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 13(8):841–847.

Xie, Y., Sha, Z., and Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology*, 1(1):9–23.

Yacobi, Y. Z., Moses, W. J., Kaganovsky, S., Sulimani, B., Leavitt, B. C., and Gitelson, A. A. (2011). NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. *Water Research*, 45(7):2428–2436.

Yacoub, M., Badran, F., and Thiria, S. (2001). A topological hierarchical clustering: Application to ocean color classification. In *Artificial Neural Networks-ICANN 2001*, pages 492–499. Springer.

Yentsch, C. S. (1960). The influence of phytoplankton pigments on the colour of sea water. *Deep Sea Research (1953)*, 7(1):1–9.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3):338–353.

Zaneveld, J. R. V. (1995). A theoretical derivation of the dependence of the remotely sensed reflectance of the ocean on the inherent optical properties. *Journal of Geophysical Research: Oceans (1978–2012)*, 100(C7):13135–13142.

Zhu, W. and Yu, Q. (2013). Inversion of chromophoric dissolved organic matter from EO-1 Hyperion imagery for turbid estuarine and coastal waters. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(6):3286–3298.

Zibordi, G., D'Alimonte, D., and Berthon, J.-F. (2004). An evaluation of depth resolution requirements for optical profiling in coastal waters. *Journal of Atmospheric and Oceanic Technology*, 21(7):1059–1073.

Zibordi, G., Mélin, F., Berthon, J.-F., and Canuti, E. (2013). Assessment of meris ocean color data products for european seas. *Ocean Science*, 9(3):521–533.

Zibordi, G. and Voss, K. J. (2010). Field radiometry and ocean color remote sensing. In *Oceanography from space*, pages 307–334. Springer.

Appendices

Appendix A

Synthetic data models

IOCCG dataset

These synthetic data originated from an existing dataset created for the International Ocean Colour Coordinating Group (IOCCG) Report 5 (IOCCG, 2006). This dataset was created with a wide dynamic range for the purpose of testing algorithms on broader scales than was possible with traditional *in situ* datasets. IOPs were generated using available optical models, while the AOPs and R_{rs} spectra were generated using Hydrolight (Mobley and Sundman, 2008) with these IOPs. Details of the IOP models that were used to create this dataset can be found in IOCCG (2006) and references therein. The data were retrieved directly from the IOCCG website; the dataset with sun at 30 degrees from zenith was used.

The entire dataset included phytoplankton concentrations of set ranges between 0.03 and 30 mg m^{-3}; however, only data from 0.03 - 5 mg m^{-3} were used, since Robertson Lain et al. (2014) found that the specific models used for the dataset did not perform as well at higher [Chl *a*]. [Chl *a*] was the set variable at 14 steps (0.03, 0.05, 0.07, 0.1, 0.15, 0.2, 0.3, 0.5, 0.7, 1, 1.5, 2, 3 and 5 mg m^{-3}) of 25 spectra each. This dataset did not include inelastic scatter such as chlorophyll fluorescence and had separate coefficients for detrital and CDOM absorption (a_d and a_g respectively); in order for these to be comparable to the *in situ* and other synthetic data, these two terms were added together to produce a_{dg}.

Case 2 modelling

The Case 2 model from Ecolight (HE5.2) (Mobley and Sundman, 2008) was chosen to augment the clustering database with synthetic spectra representing waters that might

be found at coastal sites with riverine influxes or resuspention of bottom sediments. This model provides the ability to vary the mineral and CDOM concentrations in the water to mimic *in situ* Case 2 conditions. It is also possible to change the mineral IOPs to include different types of sediments, resulting in a larger variety of spectral shapes. This dataset will hereafter be referred to as the Case 2 synthetic data. The following settings were used:

- Similar to the IOCCG dataset [Chl a] was chosen as the set variable at nine steps (0.1, 0.2, 0.5, 0.7, 1, 1.5, 2, 3 and 5 mg m^{-3}) of 15 spectra each. Mineral concentrations and CDOM absorption at 400 nm were allowed to vary randomly between 0.02 - 10 g m^{-3} and 0.01 to 0.3 m^{-1} respectively.

- Pope and Fry's *pure water* absorption and seawater scattering coefficients (Pope and Fry, 1997) were used for the pure water IOP component, as done in numerous modelling studies (e.g. Sathyendranath et al., 2001; Carder et al., 2004; Lee et al., 2010), together with an azimuthally averaged pure water phase function.

- The [Chl a] was set to be constant with depth. The specific absorption and scattering characteristics as described for the EAP model in Robertson Lain et al. (2014) were applied for the phytoplankton component. The EAP model allows a choice of phytoplankton population effective diameter; in the present study a size of 16 μ was used, which is considered a representative size for dinoflagellate assemblages, which often occur in the southern Benguela (Robertson Lain et al., 2014). Robertson Lain et al. (2014) showed that the best fit resulted from the use of a Fournier Forand phase function selected by a spectrally variable phytoplankton backscatter fraction ($b_{b\phi}/b_{\phi}$); however, the 4 component Case 2 model in HE5.2 does not allow for that option. Figure A.1 shows the phytoplankton specific $b_{b\phi}/b_{\phi}$ for the chosen effective diameter; an exponential curve, created with equation A.1, was fitted approximately to this spectra to allow the use of the power law option offered by HE5.2. It is expected that at biomass of $\leqslant$5 mg m^{-3} the scattering would be dominated by minerals and that the impact of the choice of phase function for the phytoplankton component would be minimal.

$$b_{b\phi}/b_{\phi} = 0.003\left(\frac{550}{\lambda}\right)^{-2.7} \tag{A.1}$$

- The combined CDOM and detrital absorption (a_{gd}) was set to be constant with depth; a_{gd} at 400 nm (as used in Robertson Lain et al., 2014), which is scaled in

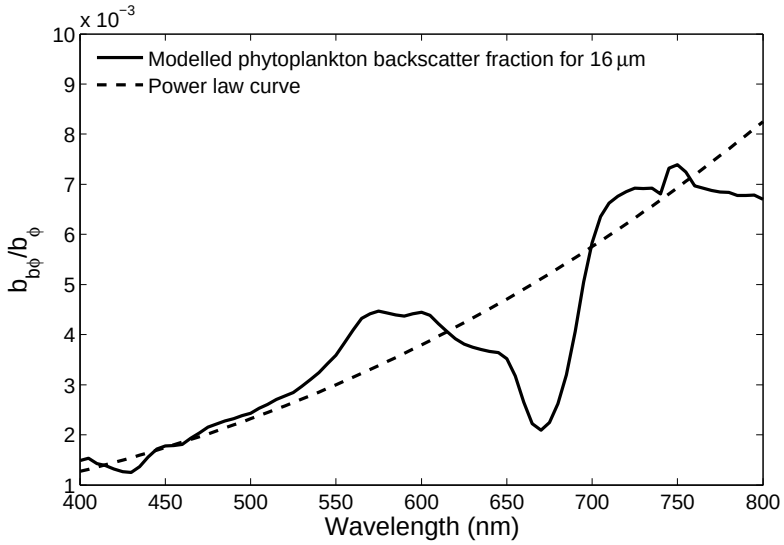


Figure A.1: The wavelength-variable phytoplankton backscatter fraction for 16 μm (solid line), which is represented by the approximate exponential curve (dashed line) in the HE5.2 Case 2 model

a non-linear fashion with [Chl a], was derived from observations in the Benguela (Bricaud et al., 1981; Roesler et al., 1989)

- Mineral concentration was set to be constant with depth. An azimuthally averaged Petzold average particle phase function (Mobley et al., 1993) was used. Mineral mass-specific absorption and scattering spectra from Ahn (1999) are provided with HE5.2, with examples for different mineral types including brown earth, calcareous sand, red clay, yellow clay and an average of all four types. In the present study only the brown earth, red clay and average sediment types were applied, since trial and error revealed that the spectral shapes imposed by the calcareous sand and yellow clay were not present in the study region. Each of the three corresponding mass-specific absorption and scattering examples were applied in turn, and 135 spectra (fifteen spectra for each of the nine [Chl a] steps) were created for each of the three types.

- Chlorophyll fluorescence quantum efficiency was set to 0.01 for all [Chl a] (Behrenfeld et al., 2009). No other inelastic scattering was applied.

- Normalized sky radiances were computed with a semi-empirical clear-sky model from Harrison and Coombes (1988) with the cloud fraction set to zero. Diffuse and direct sky irradiances were computed just above the sea surface with the RADTRAN-X model, which is the RADTAN model from Gregg and Carder (1990) extended for 300 to 1000 nm. The default values, which are considered to be adequate for the

South African marine environment, were used and included the following:

  - sea-level pressure of 101.325 kPa
  - 24 hour averaged windspeed of 5 m s^{-1}
  - horizontal visibility of 15 km
  - relative humidity of 80 %
  - precipitable water content of 2.5 cm
  - total ozone of 300 Dobson units
  - maritime airmass (airmass type 1)
  - mean earth-sun distance
  - solar zenith angle of 30 degrees

- The aerosol optical thickness at 550 nm was set to 0.261. The wind-speed wave-slope statistics model from Cox and Munk (1954a,b) was used to model the wind-blown sea surface, together with a Monte Carlo ray tracing model, as described in Mobley (1994). The index of refraction was constant for all wavelengths and set to 1.340.
- The bottom boundary was set to an infinitely deep, homogenous water body.

Appendix B

Class-specific errors and uncertainties for the combined southern Benguela and MERMAID dataset

Table B.1: The median absolute relative error (MARE) for extracted MERIS matchup data from the combined southern Benguela and MERMAID dataset. Results are shown per OWT for the following algorithms: OCMe, OC4E, OC3E, Algal2, G2B, G2B switch, G3B, G3B switch, and blended chlorophyll *a*. The *switch* shows where both the OC4E-G2B and OC4E-G3B algorithm switches, as described in section 3.2.6, were implemented. The asterisk represents the MARE value used per OWT to create the blended uncertainty product.

Class	OCMe	OC4E	OC3E	Algal2	G2B^n	G2B switch	G3B^n	G3B switch	**Blend**
1	21.9*	25.3	34.6	31.2	264	25.3	289	25.3	21.9
2	37.8*	29.9	29.0	51.1	18376	29.9	92852	29.9	35.9
3	41.8	45.4*	44.7	36.9	4905	45.4	15939	45.4	45.2
4	130	73.9*	88.0	73.4	230	70.5	585	70.5	74.2
5	47.6	27.5	33.6	51.2	210	28.1*	448	27.0	28.1
6	218	66.7	70.3	64.2	74.3	74.3	73.7*	73.7	73.7
7	166	67.7	92.9	78.8*	354	67.7	386	67.7	75.8
8	220	120	123	78.1*	133	69.9	85.9	120	77.5
9	-	-	-	[78.0]*	-	-	-	-	-
10	107	35.8	38.5	37.7*	58.4	35.8	77.5	35.8	37.7
11	131	28.7	37.4	47.2*	79.4	28.7	106	28.7	47.2
All	71.0	46.9	52.3	52.0	351	45.7	789	45.8	45.5

^{n} does not exclude negative data

* MARE value used per OWT to create the blended uncertainty product

Table B.2: The number of datapoints N used for the calculation of the MARE results in table B.1. Results are shown per OWT for the following algorithms: OCMe, OC4E, OC3E, Algal2, G2B, G2B switch, G3B, G3B switch, and blended chlorophyll *a*. The *switch* shows where both the OC4E-G2B and OC4E-G3B algorithm switches, as described in section 3.2.6, were implemented.

Class	OCMe	OC4E	OC3E	Algal2	G2B^n	G2B switch	G3B^n	G3B switch	**Blend**
1	3	3	3	3	3	3	3	3	1
2	12	12	12	12	12	12	12	12	12
3	65	65	65	65	65	65	65	65	65
4	54	54	54	52	54	54	54	54	54
5	88	88	88	88	88	88	88	88	88
6	9	9	9	9	9	9	9	9	9
7	28	28	28	28	28	28	28	28	28
8	21	21	21	21	21	21	21	21	21
9	0	0	0	0	0	0	0	0	0
10	5	5	5	5	5	5	5	5	5
11	4	4	4	4	4	4	4	4	4
All	289	289	289	287	289	289	289	289	287

^{n} does not exclude negative data

Appendix C

File names used for classification testing

Table C.1: File names for the MERIS reduced resolution data used for test images in the west coast.

MER_RR___2PRBCM20021025_080745_000000542010_00350_03410_0010.N1
MER_RR___2PRBCM20050330_083029_000000542036_00021_16107_0007.N1
MER_RR___2PNBCM20070307_081037_000000542056_00121_26227_0011.N1
MER_RR___2PRBCM20081025_081610_000000542073_00164_34787_0012.N1
MER_RR___2PRBCM20090220_080740_000000542076_00350_36476_0013.N1
MER_RR___2PRBCM20090314_081608_000000542077_00164_36791_0014.N1
MER_RR___2PRBCM20090510_082440_000000542078_00479_37607_0015.N1
MER_RR___2PRBCM20100611_084716_000000542090_00150_43290_0016.N1
MER_RR___2PRBCM20100811_083014_000000542092_00021_44163_0017.N1
MER_RR___2PRBCM20110709_084920_000000543104_00208_48929_0018.N1

Table C.2: File names for the MERIS reduced resolution data used for test images in the south coast.

MER_RR___2PNBCM20080220_081054_000000542066_00121_31237_0019.N1
MER_RR___2PRBCM20080110_075935_000000542065_00035_30650_0020.N1
MER_RR___2PRBCM20080120_074519_000000542065_00178_30793_0021.N1
MER_RR___2PRBCM20080214_075932_000000542066_00035_31151_0022.N1
MER_RR___2PRBCM20080314_074812_000000542066_00450_31566_0024.N1
MER_RR___2PRBCM20100426_075340_000000542088_00493_42631_0023.N1
MER_RR___2PRBCM20100720_082210_000000542091_00207_43848_0025.N1
MER_RR___2PRBCM20100825_075052_000000542092_00221_44363_0026.N1
MER_RR___2PRBCM20100912_082502_000000542092_00479_44621_0027.N1
MER_RR___2PRBCM20100928_082209_000000542093_00207_44850_0028.N1

Table C.3: File names for the MERIS reduced resolution data used for test images in the east coast.

MER_RR___2PNBCM20070302_072708_000000542056_00049_26155_0029.N1
MER_RR___2PRBCM20040309_072132_000000542025_00006_10581_0030.N1
MER_RR___2PRBCM20040527_073835_000000542027_00135_11712_0031.N1
MER_RR___2PRBCM20070425_072955_000000542057_00321_26928_0032.N1
MER_RR___2PRBCM20080218_073241_000000542066_00092_31208_0033.N1
MER_RR___2PRBCM20080324_073244_000000542067_00092_31709_0034.N1
MER_RR___2PRBCM20080409_072954_000000542067_00321_31938_0035.N1
MER_RR___2PRBCM20080425_072701_000000542068_00049_32167_0036.N1
MER_RR___2PRBCM20080605_073823_000000542069_00135_32754_0037.N1
MER_RR___2PRBCM20090416_073820_000000542078_00135_37263_0038.N1

Table C.4: File names for the MERIS reduced resolution data used for satellite data extraction.

MER_RR___2PNACR20070225_075549_000026042055_00479_26084_0000.N1
MER_RR___2PNACR20070307_074009_000026082056_00121_26227_0000.N1
MER_RR___2PNACR20070313_075117_000025862056_00207_26313_0000.N1
MER_RR___2PRACR20021025_074318_000024012010_00350_03410_0000.N1
MER_RR___2PRACR20050330_075718_000026212036_00021_16107_0000.N1
MER_RR___2PRACR20050405_080807_000026252036_00107_16193_0000.N1
MER_RR___2PRACR20080822_075151_000026302071_00250_33871_0000.N1
MER_RR___2PRACR20080904_074442_000026292071_00436_34057_0000.N1
MER_RR___2PRACR20081006_074257_000026322072_00393_34515_0000.N1
MER_RR___2PRACR20081025_074801_000026342073_00164_34787_0000.N1
MER_RR___2PRACR20090223_074425_000026042076_00393_36519_0000.N1
MER_RR___2PRACR20090310_081113_000026102077_00107_36734_0000.N1
MER_RR___2PRACR20090314_074450_000026112077_00164_36791_0000.N1
MER_RR___2PRACR20090411_080126_000026282078_00064_37192_0000.N1
MER_RR___2PRACR20090414_080651_000026292078_00107_37235_0000.N1
MER_RR___2PRACR20090501_073042_000026372078_00350_37478_0000.N1
MER_RR___2PRACR20090503_080754_000026382078_00379_37507_0000.N1
MER_RR___2PRACR20090510_074713_000026402078_00479_37607_0000.N1
MER_RR___2PRACR20100425_074824_000026352088_00479_42617_0000.N1
MER_RR___2PRACR20100530_074550_000026432089_00479_43118_0000.N1
MER_RR___2PRACR20100611_080824_000026432090_00150_43290_0000.N1
MER_RR___2PRACR20100811_075326_000026312092_00021_44163_0000.N1
MER_RR___2PRACR20100912_075115_000026292092_00479_44621_0000.N1
MER_RR___2PRACR20110709_081048_000026333104_00208_48929_0000.N1

Appendix D

Cluster means and covariance matrices

Table D.1: Mean R_{rs} (in sr^{-1}) spectra for the eleven optical water types (OWTs) at the nine MERIS bands

OWT	$R_{rs}442.5$	$R_{rs}490$	$R_{rs}510$	$R_{rs}560$	$R_{rs}620$	$R_{rs}665$	$R_{rs}681.25$	$R_{rs}708.75$	R_{rs} 753.75
1	0.0105	0.0078	0.0049	0.0024	0.0006	0.0003	0.0003	0.0002	0.0001
2	0.0071	0.0061	0.0042	0.0022	0.0005	0.0003	0.0003	0.0002	0.0001
3	0.0037	0.0040	0.0034	0.0025	0.0006	0.0004	0.0005	0.0002	0.0001
4	0.0036	0.0053	0.0056	0.0064	0.0021	0.0013	0.0014	0.0008	0.0003
5	0.0012	0.0019	0.0022	0.0030	0.0011	0.0008	0.0013	0.0007	0.0003
6	0.0023	0.0024	0.0026	0.0049	0.0035	0.0021	0.0028	0.0043	0.0013
7	0.0041	0.0066	0.0072	0.0100	0.0043	0.0025	0.0026	0.0021	0.0006
8	0.0060	0.0105	0.0111	0.0122	0.0042	0.0027	0.0025	0.0015	0.0004
9	0.0061	0.0094	0.0108	0.0165	0.0111	0.0070	0.0066	0.0041	0.0012
10	0.0099	0.0151	0.0162	0.0200	0.0089	0.0054	0.0049	0.0028	0.0008
11	0.0116	0.0177	0.0198	0.0266	0.0162	0.0108	0.0099	0.0063	0.0018

Table D.2: Covariance matrices for optical water types 1 and 2

OWT1								
3.25E-06	1.70E-06	6.02E-07	2.11E-07	3.04E-08	-9.58E-09	-2.12E-08	-1.11E-08	-1.17E-08
1.70E-06	1.85E-06	1.19E-06	7.23E-07	1.36E-07	5.91E-08	4.46E-08	2.24E-08	2.96E-09
6.02E-07	1.19E-06	1.02E-06	6.70E-07	1.50E-07	8.18E-08	6.99E-08	3.83E-08	9.61E-09
2.11E-07	7.23E-07	6.70E-07	4.80E-07	1.08E-07	5.79E-08	4.95E-08	2.69E-08	6.37E-09
3.04E-08	1.36E-07	1.50E-07	1.08E-07	3.24E-08	1.95E-08	1.67E-08	1.05E-08	2.42E-09
-9.58E-09	5.91E-08	8.18E-08	5.79E-08	1.95E-08	1.34E-08	1.18E-08	7.51E-09	2.14E-09
-2.12E-08	4.46E-08	6.99E-08	4.95E-08	1.67E-08	1.18E-08	1.10E-08	6.82E-09	2.08E-09
-1.11E-08	2.24E-08	3.83E-08	2.69E-08	1.05E-08	7.51E-09	6.82E-09	5.54E-09	1.58E-09
-1.17E-08	2.96E-09	9.61E-09	6.37E-09	2.42E-09	2.14E-09	2.08E-09	1.58E-09	1.03E-09
OWT2								
1.21E-06	7.75E-07	3.24E-07	3.29E-08	8.09E-09	-1.35E-08	-4.67E-08	-7.71E-09	-1.03E-08
7.75E-07	9.26E-07	6.54E-07	4.04E-07	9.06E-08	3.38E-08	9.09E-09	1.52E-08	1.34E-08
3.24E-07	6.54E-07	6.88E-07	6.01E-07	1.61E-07	9.51E-08	9.82E-08	5.03E-08	2.69E-08
3.29E-08	4.04E-07	6.01E-07	6.67E-07	1.95E-07	1.24E-07	1.48E-07	6.69E-08	2.89E-08
8.09E-09	9.06E-08	1.61E-07	1.95E-07	7.30E-08	4.93E-08	5.76E-08	2.90E-08	1.39E-08
-1.35E-08	3.38E-08	9.51E-08	1.24E-07	4.93E-08	3.78E-08	4.57E-08	2.28E-08	1.21E-08
-4.67E-08	9.09E-09	9.82E-08	1.48E-07	5.76E-08	4.57E-08	6.28E-08	2.77E-08	1.51E-08
-7.71E-09	1.52E-08	5.03E-08	6.69E-08	2.90E-08	2.28E-08	2.77E-08	1.52E-08	8.19E-09
-1.03E-08	1.34E-08	2.69E-08	2.89E-08	1.39E-08	1.21E-08	1.51E-08	8.19E-09	1.47E-08

Table D.3: Covariance matrices for optical water types 3 and 4

OWT3								
1.15E-06	9.44E-07	5.27E-07	-3.03E-07	-2.82E-07	-1.90E-07	-2.30E-07	-1.42E-07	-2.64E-08
9.44E-07	1.70E-06	1.30E-06	3.25E-08	-3.99E-07	-2.75E-07	-4.24E-07	-2.53E-07	-1.03E-07
5.27E-07	1.30E-06	1.20E-06	5.16E-07	-1.03E-07	-8.59E-08	-2.14E-07	-1.24E-07	-6.19E-08
-3.03E-07	3.25E-08	5.16E-07	1.87E-06	8.87E-07	5.17E-07	5.13E-07	3.22E-07	8.29E-08
-2.82E-07	-3.99E-07	-1.03E-07	8.87E-07	6.22E-07	3.83E-07	4.25E-07	2.70E-07	9.27E-08
-1.90E-07	-2.75E-07	-8.59E-08	5.17E-07	3.83E-07	2.46E-07	2.83E-07	1.76E-07	6.57E-08
-2.30E-07	-4.24E-07	-2.14E-07	5.13E-07	4.25E-07	2.83E-07	3.70E-07	2.18E-07	8.68E-08
-1.42E-07	-2.53E-07	-1.24E-07	3.22E-07	2.70E-07	1.76E-07	2.18E-07	1.40E-07	5.26E-08
-2.64E-08	-1.03E-07	-6.19E-08	8.29E-08	9.27E-08	6.57E-08	8.68E-08	5.26E-08	2.98E-08
OWT4								
3.48E-07	1.90E-07	4.66E-08	-2.38E-07	-1.43E-07	-9.21E-08	-1.07E-07	-6.39E-08	-1.85E-08
1.90E-07	4.93E-07	4.25E-07	3.49E-08	-7.56E-08	-6.31E-08	-1.41E-07	-6.38E-08	-2.54E-08
4.66E-08	4.25E-07	4.93E-07	3.72E-07	8.35E-08	4.17E-08	-2.28E-08	-1.93E-10	-3.41E-09
-2.38E-07	3.49E-08	3.72E-07	1.41E-06	5.92E-07	3.62E-07	3.83E-07	2.34E-07	7.13E-08
-1.43E-07	-7.56E-08	8.35E-08	5.92E-07	3.41E-07	2.16E-07	2.38E-07	1.52E-07	5.44E-08
-9.21E-08	-6.31E-08	4.17E-08	3.62E-07	2.16E-07	1.47E-07	1.78E-07	1.03E-07	4.05E-08
-1.07E-07	-1.41E-07	-2.28E-08	3.83E-07	2.38E-07	1.78E-07	2.68E-07	1.36E-07	5.69E-08
-6.39E-08	-6.38E-08	-1.93E-10	2.34E-07	1.52E-07	1.03E-07	1.36E-07	8.36E-08	3.29E-08
-1.85E-08	-2.54E-08	-3.41E-09	7.13E-08	5.44E-08	4.05E-08	5.69E-08	3.29E-08	1.73E-08

Table D.4: Covariance matrices for optical water types 5 and 6

OWT5								
2.74E-07	2.29E-07	1.70E-07	2.14E-08	-2.23E-08	-1.72E-08	-8.03E-08	-7.74E-08	-2.03E-08
2.29E-07	3.37E-07	2.91E-07	4.97E-08	-9.06E-08	-5.73E-08	-1.63E-07	-1.72E-07	-6.02E-08
1.70E-07	2.91E-07	2.88E-07	1.35E-07	-5.00E-08	-2.76E-08	-1.08E-07	-1.46E-07	-5.13E-08
2.14E-08	4.97E-08	1.35E-07	4.74E-07	2.37E-07	1.53E-07	2.32E-07	1.30E-07	5.35E-08
-2.23E-08	-9.06E-08	-5.00E-08	2.37E-07	2.48E-07	1.48E-07	2.41E-07	2.31E-07	8.90E-08
-1.72E-08	-5.73E-08	-2.76E-08	1.53E-07	1.48E-07	1.03E-07	1.80E-07	1.27E-07	5.25E-08
-8.03E-08	-1.63E-07	-1.08E-07	2.32E-07	2.41E-07	1.80E-07	3.75E-07	2.48E-07	1.02E-07
-7.74E-08	-1.72E-07	-1.46E-07	1.30E-07	2.31E-07	1.27E-07	2.48E-07	3.34E-07	1.19E-07
-2.03E-08	-6.02E-08	-5.13E-08	5.35E-08	8.90E-08	5.25E-08	1.02E-07	1.19E-07	4.89E-08
OWT6								
5.27E-07	3.95E-07	3.24E-07	2.72E-07	2.04E-07	1.35E-07	6.57E-08	7.62E-08	3.19E-08
3.95E-07	5.31E-07	5.66E-07	7.02E-07	3.27E-07	2.09E-07	4.19E-08	-3.71E-07	-8.39E-08
3.24E-07	5.66E-07	6.70E-07	9.81E-07	4.15E-07	2.51E-07	8.06E-08	-5.04E-07	-1.23E-07
2.72E-07	7.02E-07	9.81E-07	2.09E-06	8.38E-07	4.08E-07	2.66E-07	-7.11E-07	-2.06E-07
2.04E-07	3.27E-07	4.15E-07	8.38E-07	6.54E-07	3.24E-07	2.95E-07	4.33E-07	1.26E-07
1.35E-07	2.09E-07	2.51E-07	4.08E-07	3.24E-07	1.83E-07	1.51E-07	2.13E-07	6.98E-08
6.57E-08	4.19E-08	8.06E-08	2.66E-07	2.95E-07	1.51E-07	2.74E-07	5.49E-07	1.57E-07
7.62E-08	-3.71E-07	-5.04E-07	-7.11E-07	4.33E-07	2.13E-07	5.49E-07	3.43E-06	9.37E-07
3.19E-08	-8.39E-08	-1.23E-07	-2.06E-07	1.26E-07	6.98E-08	1.57E-07	9.37E-07	2.68E-07

Table D.5: Covariance matrices for optical water types 7 and 8

OWT7								
1.86E-06	1.45E-06	7.19E-07	-9.95E-07	-8.71E-07	-4.79E-07	-5.94E-07	-6.85E-07	-1.86E-07
1.45E-06	3.47E-06	3.02E-06	-8.24E-07	-1.85E-06	-8.46E-07	-1.22E-06	-2.01E-06	-6.04E-07
7.19E-07	3.02E-06	3.10E-06	1.50E-07	-1.27E-06	-4.77E-07	-8.36E-07	-1.83E-06	-5.49E-07
-9.95E-07	-8.24E-07	1.50E-07	2.67E-06	2.02E-06	1.16E-06	1.19E-06	7.93E-07	2.32E-07
-8.71E-07	-1.85E-06	-1.27E-06	2.02E-06	3.12E-06	1.66E-06	1.75E-06	1.98E-06	5.87E-07
-4.79E-07	-8.46E-07	-4.77E-07	1.16E-06	1.66E-06	9.23E-07	9.55E-07	9.83E-07	2.94E-07
-5.94E-07	-1.22E-06	-8.36E-07	1.19E-06	1.75E-06	9.55E-07	1.07E-06	1.22E-06	3.67E-07
-6.85E-07	-2.01E-06	-1.83E-06	7.93E-07	1.98E-06	9.83E-07	1.22E-06	2.30E-06	6.54E-07
-1.86E-07	-6.04E-07	-5.49E-07	2.32E-07	5.87E-07	2.94E-07	3.67E-07	6.54E-07	2.06E-07
OWT8								
3.59E-06	1.86E-06	5.19E-07	-1.04E-06	-8.57E-07	-6.68E-07	-6.63E-07	-4.95E-07	-5.07E-08
1.86E-06	2.48E-06	1.84E-06	1.54E-07	-5.82E-07	-3.34E-07	-3.46E-07	-2.56E-07	-5.02E-08
5.19E-07	1.84E-06	2.02E-06	1.68E-06	3.46E-07	2.83E-07	2.52E-07	1.39E-07	2.43E-08
-1.04E-06	1.54E-07	1.68E-06	5.19E-06	2.86E-06	1.75E-06	1.66E-06	1.03E-06	2.34E-07
-8.57E-07	-5.82E-07	3.46E-07	2.86E-06	2.30E-06	1.46E-06	1.36E-06	8.68E-07	2.19E-07
-6.68E-07	-3.34E-07	2.83E-07	1.75E-06	1.46E-06	9.87E-07	9.26E-07	6.02E-07	1.52E-07
-6.63E-07	-3.46E-07	2.52E-07	1.66E-06	1.36E-06	9.26E-07	8.88E-07	5.81E-07	1.48E-07
-4.95E-07	-2.56E-07	1.39E-07	1.03E-06	8.68E-07	6.02E-07	5.81E-07	3.98E-07	9.78E-08
-5.07E-08	-5.02E-08	2.43E-08	2.34E-07	2.19E-07	1.52E-07	1.48E-07	9.78E-08	3.22E-08

Table D.6: Covariance matrices for optical water types 9 and 10

OWT9								
9.66E-07	9.47E-07	9.39E-07	1.09E-06	2.10E-07	8.37E-08	-6.28E-08	-3.71E-07	-9.40E-08
9.47E-07	2.36E-06	2.61E-06	1.95E-06	-3.13E-07	5.63E-08	-1.39E-07	-3.14E-07	-1.61E-07
9.39E-07	2.61E-06	2.99E-06	2.42E-06	-1.37E-07	2.49E-07	4.03E-08	-1.61E-07	-9.66E-08
1.09E-06	1.95E-06	2.42E-06	3.52E-06	1.59E-06	1.17E-06	9.88E-07	5.74E-07	2.89E-07
2.10E-07	-3.13E-07	-1.37E-07	1.59E-06	4.07E-06	2.84E-06	2.52E-06	1.51E-06	5.29E-07
8.37E-08	5.63E-08	2.49E-07	1.17E-06	2.84E-06	2.41E-06	2.24E-06	1.21E-06	4.23E-07
-6.28E-08	-1.39E-07	4.03E-08	9.88E-07	2.52E-06	2.24E-06	2.24E-06	1.35E-06	4.79E-07
-3.71E-07	-3.14E-07	-1.61E-07	5.74E-07	1.51E-06	1.21E-06	1.35E-06	1.31E-06	4.21E-07
-9.40E-08	-1.61E-07	-9.66E-08	2.89E-07	5.29E-07	4.23E-07	4.79E-07	4.21E-07	1.72E-07
OWT10								
4.33E-06	2.49E-06	1.57E-06	9.74E-07	-5.29E-07	-7.68E-07	-8.40E-07	-7.95E-07	-1.51E-07
2.49E-06	3.78E-06	3.51E-06	1.67E-06	-7.82E-07	-4.69E-07	-4.38E-07	-3.56E-07	-1.34E-07
1.57E-06	3.51E-06	3.71E-06	2.56E-06	2.87E-07	3.52E-07	3.46E-07	1.75E-07	-4.08E-09
9.74E-07	1.67E-06	2.56E-06	5.22E-06	4.42E-06	2.82E-06	2.54E-06	1.40E-06	4.01E-07
-5.29E-07	-7.82E-07	2.87E-07	4.42E-06	5.71E-06	3.69E-06	3.33E-06	1.98E-06	5.92E-07
-7.68E-07	-4.69E-07	3.52E-07	2.82E-06	3.69E-06	2.52E-06	2.31E-06	1.44E-06	4.09E-07
-8.40E-07	-4.38E-07	3.46E-07	2.54E-06	3.33E-06	2.31E-06	2.13E-06	1.35E-06	3.78E-07
-7.95E-07	-3.56E-07	1.75E-07	1.40E-06	1.98E-06	1.44E-06	1.35E-06	9.11E-07	2.46E-07
-1.51E-07	-1.34E-07	-4.08E-09	4.01E-07	5.92E-07	4.09E-07	3.78E-07	2.46E-07	7.12E-08

Table D.7: Covariance matrix for optical water type 11

OWT11							
4.36E-06	3.22E-06	2.59E-06	1.83E-06	-3.78E-07	-7.84E-07	-8.71E-07	-1.73E-06	-2.54E-07
3.22E-06	7.14E-06	8.09E-06	7.64E-06	4.63E-06	4.66E-06	4.59E-06	4.16E-06	1.25E-06
2.59E-06	8.09E-06	9.64E-06	9.50E-06	6.62E-06	6.77E-06	6.70E-06	6.43E-06	1.84E-06
1.83E-06	7.64E-06	9.50E-06	1.04E-05	8.77E-06	8.59E-06	8.46E-06	8.22E-06	2.35E-06
-3.78E-07	4.63E-06	6.62E-06	8.77E-06	1.08E-05	1.05E-05	1.03E-05	9.95E-06	2.94E-06
-7.84E-07	4.66E-06	6.77E-06	8.59E-06	1.05E-05	1.05E-05	1.04E-05	1.01E-05	2.95E-06
-8.71E-07	4.59E-06	6.70E-06	8.46E-06	1.03E-05	1.04E-05	1.03E-05	1.00E-05	2.92E-06
-1.73E-06	4.16E-06	6.43E-06	8.22E-06	9.95E-06	1.01E-05	1.00E-05	1.05E-05	2.94E-06
-2.54E-07	1.25E-06	1.84E-06	2.35E-06	2.94E-06	2.95E-06	2.92E-06	2.94E-06	8.81E-07